Universidad
Carlos III de Madrid
www.uc3m.es

# TESIS DOCTORAL

# Discovering common features in a large set of disaggregates. Methodology, modeling and forecasting

**Autor:**

**Guillermo Carlomagno**

**Director:**

**Antoni Espasa**

**Departamento de Estadística**

Getafe, Febrero 2016

*This page intentionally left blank.*

# Agradecimientos

En primer lugar quiero agradecer y dedicar este trabajo a mis padres. Ellos me transmitieron el deseo por aprender y descubrir, de ellos aprendí a cuestionarlo todo, a ponerme objetivos desafiantes y a no bajar nunca los brazos. Esas lecciones fueron fundamentales en este proyecto. Gracias a ellos también por su apoyo incondicional.

A Gabi, por acompañarnos siempre. Por haber decidido recorrer este camino juntos. Por soportarme siempre.

Un agradecimiento muy especial a mi maestro, Toni. Por las infinitas conversaciones, he aprendido más de ellas que de la lectura de los más de cien artículos que se citan en esta tesis. Gracias por tu tiempo, tu dedicación, por tu exigencia, por empujarme a mover mis límites. Por hacerme saber que contaba contigo.

Un agradecimiento también especial a mi segundo maestro, Fernando. Sin dudas, el principal responsable de que haya elegido recorrer este camino. El me lo mostró y fue mi referencia. Su apoyo y su confianza en mi fueron fundamentales en la etapa previa a la tesis.

A mis hermanos, a mis amigos y a mis familiares, con los que cuento siempre. Entre mis amigos, un agradecimiento especial a Daniel Egger no solo por los innumerables momentos de amistad, también por haberse leído una versión anterior de esta tesis y ayudarme a mejorarla.

Quiero agradecer también a la Universidad Carlos III y al Departamento de Estadística por el apoyo material que me brindaron durante estos cinco años.

*This page intentionally left blank.*

*Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.*

— Albert Einstein

# Abstract

Macroeconomic variables are, in general, weighted averages of a large number of components. Therefore, the usual focus on the aggregate alone implies neglecting a large amount of information. The general objective of this thesis is to develop a procedure to model and forecast all the components of a macro or business variable at the maximum level of disaggregation. We denote these components as *basic components* and usually their number is large (hundreds). Our strategy consist of identifying and estimating relevant relationships between the basic components and then exploiting those relationships in single-equation models for all the disaggregates. This strategy can produce relatively precise forecasts of the components and may lead to an accurate indirect forecast for the aggregate.

Since we are interested in modeling and forecasting all the disaggregates, a large estimation effort is inevitable and the advantages of disaggregation could be off-set by estimation uncertainty issues. Our approach for dealing with the informational losses vs. estimation uncertainty trade-off is the consideration of common features, as proposed by Espasa and Mayo-Burgos (2013). Their suggestion is to try to discover blocks of components that share unique common features (trends and cycles), and then to include the restrictions implied by those commonalities in single-equation models for the components. The search for those blocks is carried out by performing common features tests between all the $N(N-1)/2$ pairs that exist in a set of $N$ components. As the authors argue, the level of disaggregation used in this procedure must be the maximum available since *ad-hoc* sub-aggregates may add up series that do not share common features.

An important aim of this thesis is to show that the mentioned pairwise strategy can be used as an objective method to discover blocks of components sharing single common features. That is, blocks of components can be constructed by first testing for common features between all the $N(N-1)/2$ pairs of series and then grouping those that share a unique common feature.

In this thesis we study by analytical and simulation methods the statistical properties of the procedure, thus, providing a solid base for its application.

An important theoretical result is that the pairwise cointegration tests in a block of series that share a unique common trend are asymptotically equivalent, in the sense that the probability that all tests deliver the same conclusion tends to 1 as $T$ goes to infinity, independently of the number of series. Thus, in this case the multiple testing is not an issue.

In a Monte Carlo experiment, we confirm the asymptotic results and compare the performance of the pairwise approach with that of a Dynamic Factor Models (DFM) alternative. Additionally, we extend the procedure to make it robust when applied to real data which may be subject to irregularities and/or to short samples problems.

In regards to common cycles tests, we analytically justify the validity of proceeding in a pairwise fashion for discovering 'single-cycle' subsets. In contrast with our results on common trends, we show that the multiple testing problem is in fact present in common cycles tests. This issue makes the probability of including a large proportion of the true series inside the estimated 'single-cycle' set to be a decreasing function of the true size of the subset. This is an undesirable property, for we want our procedure to work well not only when subsets are small, but also when they are large. We show that a simple 'relaxation' strategy solves the problem, with almost zero cost.

The application of the pairwise approach to the US CPI and the US IPI lead to interesting conclusions. First, a detailed analysis of outliers indicate that while the 'great moderation' period can be characterized by a negative significant mean shift in the proportion of components with outliers, the sub-prime crises is characterized by a significant and positive mean shift. Additionally, we found that the *aggregated outlier* — a series constructed by aggregating the outliers of the components — helps to improve the model of the aggregate. This is because there are some outliers that are not identifiable in the aggregated series. Finally, in pseudo out of sample forecasting exercises, we compare the ability of the pairwise approach to forecast the aggregate with other indirect and direct procedures. The general conclusions are similar for both applications; the consideration of common features' restrictions helps to improve indirect forecasts and to beat direct procedures.

# Resumen

Las variables macroeconómicas son, en general, promedios ponderados de un gran número de componentes. Por tanto, el enfoque usual, que considera únicamente agregado implica descartar una gran cantidad de información. El objetivo general de esta tesis es desarrollar un procedimiento para modelizar y predecir todos los componentes de una variable macro-económica al máximo nivel de desagregación. Denotamos estos componentes como *componentes básicos* y por lo general su número es grande (cientos). Nuestra estrategia consiste en identificar y estimar relaciones relevantes entre los componentes básicos y luego explotar esas relaciones en modelos uni-ecuacionales para todos los componentes. Esta estrategia puede generar predicciones relativamente precisas de los componentes y puede dar lugar a una predicción indirecta certera para el agregado.

Dado que estamos interesados en modelizar y predecir todos los componentes (no solamente el agregado), un gran esfuerzo de estimación es inevitable y las ventajas de la desagregación podrían perderse por cuestiones de incertidumbre de estimación. Nuestro enfoque para hacer frente al trade-off de pérdidas de información vs. incertidumbre en la estimación es la consideración de características comunes, como proponen Espasa and Mayo-Burgos (2013). Su sugerencia consiste en tratar de descubrir bloques de componentes que compartan características comunes únicas — *common features* — (tendencias y ciclos), y luego incluir las restricciones que se derivan de dichas características comunes en modelos uni-ecuacionales para todos los componentes. La búsqueda de esos bloques se lleva a cabo mediante la realización de contrastes de características comunes entre todos los $N(N-1)/2$ pares que existen en un conjunto de $N$ componentes. Como los autores argumentan, el nivel de desagregación utilizado en este procedimiento debe ser el máximo disponible ya que sub-agregados *ad-hoc* pueden agregar componentes que no comparten características comunes.

Un objetivo importante de esta tesis es demostrar que la estrategia de pares mencionada

puede ser utilizada como un método objetivo para descubrir bloques de componentes que comparten características comunes individuales. Es decir, los bloques de componentes pueden ser construidos contrastando características comunes entre todos los $N(N-1)/2$ pares de series y, a continuación agrupando las que comparten una característica común única. En esta tesis se estudia mediante métodos analíticos y de simulación las propiedades estadísticas del procedimiento, proporcionando, por tanto, una base sólida para su aplicación.

Un propósito específico central de esta tesis es demostrar que la estrategia por pares propuesta inicialmente por Espasa and Mayo-Burgos (2013) puede ser utilizada como un método objetivo para descubrir bloques de componentes que comparten características comunes individuales. Es decir, los bloques de componentes se pueden construir mediante la búsqueda de características comunes entre todos los $N(N-1)/2$ pares de series y luego agrupar los que comparten una característica común única. Nuestra contribución principal en este aspecto consiste en proporcionar las propiedades estadísticas de la estrategia utilizando procedimientos analíticos y de Monte Carlo.

Un resultado teórico importante es que los contrastes de cointegración por pares dentro de un bloque de series que comparten una única tendencia común asintóticamente equivalentes, en el sentido de que la probabilidad de que todos los contrastes arrojen la misma conclusión es tiende a 1 cuando $T$ tiende a infinito, independientemente del número de series. Por lo tanto, en este caso, no hay un problema de contrastes múltiples.

En un experimento de Monte Carlo, confirmamos los resultados asintóticos y comparamos el desempeño del enfoque por pares con el de una alternativa de Dynamic Factor Models (DFM). Adicionalmente extendemos el procedimiento para que sea robusto cuando se aplica a datos reales que pueden estar sujetos a irregularidades y/o problemas de muestras cortas.

En lo que respecta a las pruebas de ciclos comunes, justificamos analíticamente la validez del procedimiento por pares para descubrir subconjuntos de "un sólo ciclo común". En contraste con nuestros resultados para tendencias comunes, demostramos que el problema de contrastes múltiples sí está presente en las pruebas de ciclos comunes. Este problema hace que la probabilidad de incluir una gran proporción de series correctas dentro del subconjunto estimado sea una función decreciente del verdadero tamaño del subconjunto. Esta es una propiedad indeseable, porque queremos que nuestro procedimiento funcione bien no sólo cuando los subconjuntos son pequeños, sino también cuando son grandes. Demostramos sin embargo que una estrategia

simple “relajación” resuelve el problema, casi con cero costo.

La aplicación del enfoque por parejas al IPC y al IPI de Estados Unidos conduce a conclusiones interesantes. En primer lugar, el análisis detallado de valores atípicos indica que mientras el período de “la gran moderación” puede ser caracterizado por un cambio negativo y significativo en la media de la proporción de componentes con valores atípicos, las crisis sub-prime se caracteriza por un cambio de media significativo y positivo. Además, encontramos que el *valor atípico agregado* —una serie construida mediante la agregación de los valores atípicos de los componentes — ayuda a mejorar el modelo del agregado. Esto se debe a que algunos valores atípicos no son identificables en la serie agregada. Finalmente, en ejercicios de predicción seudo furera de muestra, comparamos la capacidad del enfoque por parejas para predecir el agregado con otros procedimientos indirectos y directos. Las conclusiones generales son similares en ambas aplicaciones; la consideración de restricciones derivadas de las características comunes ayuda a mejorar las predicciones indirectas y a superar a los procedimientos directos.

# Contents

# List of Figures

# List of Tables

*This page intentionally left blank.*

# Chapter I

# Introduction

Macroeconomic variables are, in general, weighted averages of a large number of components. Therefore, the usual focus on the aggregate alone implies neglecting a large amount of information. The general objective of this thesis is to develop a procedure to model and forecast all the components of a macro or business variable at the maximum level of sectoral disaggregation. We denote these components as *basic components* and usually their number is large (hundreds). Our strategy consist of identifying and estimating relevant relationships between the basic components and then exploiting those relationships in single-equation models for all the disaggregates. This strategy can produce relatively precise forecasts of the components and may lead to an accurate indirect forecast for the aggregate.

Given this objective, the first question that arises is why disaggregation is relevant. There are at least four reasons for disaggregating a macroeconomic or business variable. The most important one is that the analysis of disaggregated data by itself may be of interest for decision makers. Consider for example the case of central banks, who take into account the expected inflation to take monetary policy decisions. An inflation forecast close to the target may trigger different monetary policy decisions if all the components are expected to grow at a similar rate or if, say, core inflation is above the target and non-core components below it. For this reason, splitting the CPI in core and non-core components has became a widespread practice. Nonetheless, this level of disaggregation is still too limited. It could be the case that the forecasts of discordant inflationary components are limited just to a specific group of goods or services, and ignoring this fact could not help to a proper assessment of inflation.

In addition to improving the information for monetary policy decisions, disaggregating the

CPI and providing forecasts for the components could be very useful for decision makers in the different sectors of the economy. Disaggregation also allows to study the interrelations between sectoral prices. Understanding how prices in one industry are related to those of others is relevant not only for making decisions at the firm level but also for an economic policy point of view. For example, taxation changes in one sector may affect the price level of that sector but may also have spillover effects on other ones. A disaggregated analysis that studies the interconnection between sectoral prices would be useful for anticipating those spillovers.

When the interest is in the disaggregates, the use of the aggregate as a 'proxy' of the component(s) of interest can be completely misleading. Figure I.1 and figure I.2 show the aggregated Consumer Price Index, and Industrial Production Index for the US economy (*US CPI* and *US IPI*, respectively) and all their components. An evident observation of the two plots is that the weighted averages hide a great heterogeneity among the components (note the difference in the scale between the panels of the plots). This heterogeneity has two important implications; first, the analysis of the aggregate may be irrelevant when the interest is in the components; second, even when the interest is in the 'big-picture', disregarding the components implies a great deal of informational losses, and should not be the 'default' option.

A second reason for disaggregating comes when acknowledging the importance of understanding the micro heterogeneity and the statistical effects of the aggregation process in order to explain the dynamic properties of aggregated variables. Due to its crucial implications on the effects of monetary policy, persistence has been one of the most studied dynamic properties of price (inflation) series. Traditional analysis on aggregated series agree in classifying inflation as highly persistent, giving rise to the *sticky price* macroeconomic models. Nominal rigidities are nowadays a central feature of price setting and general macroeconomic models constructed for analyzing the business cycle. However, recent evidence on disaggregated series seems to question the nominal rigidity assumption. There is strong empirical evidence on low persistence, high frequency changes and high volatility of disaggregated prices.

There are three types of explanations regarding the seeming contradiction between aggregated and disaggregated evidence on price stickiness. In the first one, the difference is due to the aggregation bias in heterogeneous ARIMA models. In the second, the difference is due to the heterogeneous response of sectoral and aggregated series to macroeconomic shocks. Finally, some authors argue that structural breaks are the cause of aggregated stickiness. For a detailed

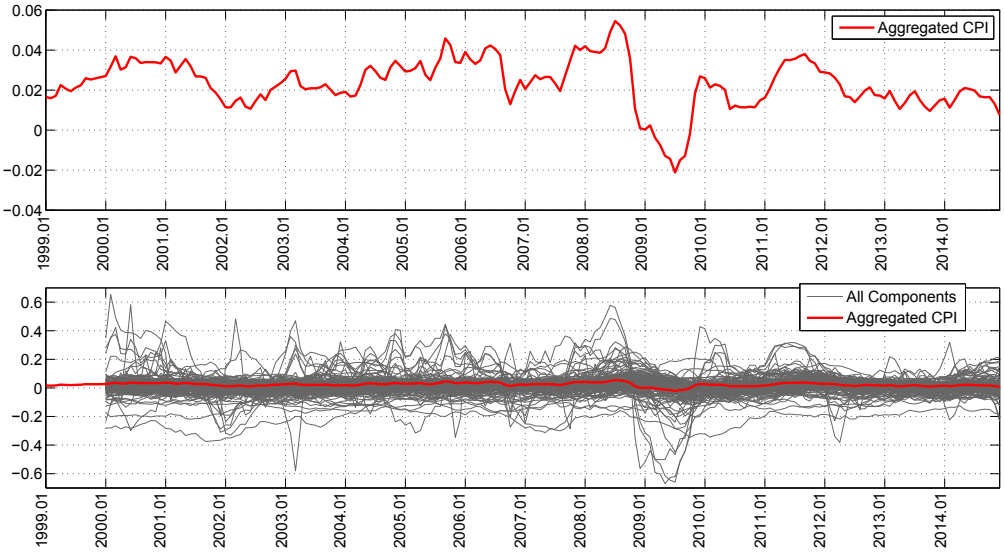



Figure I.1: $\Delta_{12}(log\ US\ CPI)$ and its 164 components, 1999.1 - 2014.12

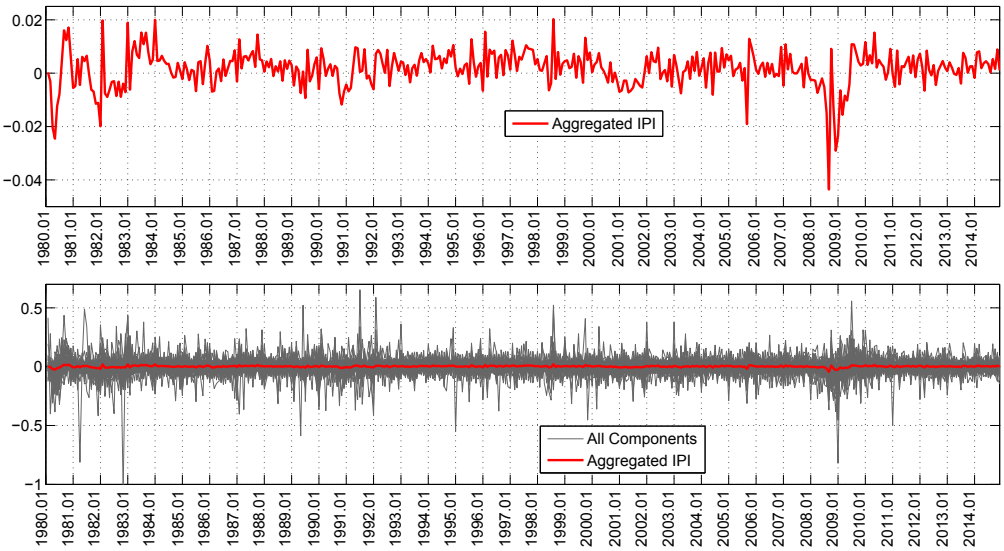



Figure I.2: $\Delta(log\ US\ IPI)$ and its 134 components, 1980.1 - 2014.12 (seasonally adjusted)

discussion see, e.g., Bils and Klenow (2004), Lunnemann and Mathä (2004), Imbs et al. (2005), Clark (2006), Altissimo et al. (2007), Boivin et al. (2009), and Beck et al. (2015)). In any case, it seems evident that disaggregated evidence has much to say in the discussion of price stickiness.

The third reason for disaggregating a macro-variable is that disaggregation is relevant for the comparative analysis of the components, what is clearly important when dealing with prices because relative prices are a key indicator of the economy. There are two types of theories that highlight the importance of relative prices; those linking relative prices variability with aggregated inflation and ascribing general welfare effects either to this relationship or to the

variability of the relative prices *per se.* (see, Keynes (1924); Hayek (1945); Alchian (1969); Blejer and Leiderman (1980); Fischer et al. (1981); and Balke and Wynne (2000)); and those theories that predict some feature of the long term dynamic properties of certain relative prices (One Price Law — and its associated theories — and the Presbisch-Singer hypothesis; see Hagen (1989) and Harvey et al. (2010)). In both types of theories, the disaggregated analysis plays a central role.

The fourth reason for disaggregating macro-variables is that the use of both, disaggregated information and disaggregated forecasts may lead to better forecasts of the aggregate (see, e.g., Espasa et al. (2002), Giacomini and Granger (2004), Hendry and Hubrich (2005, 2011), Espasa and Albacete (2007), Castle and Hendry (2010), Espasa and Mayo-Burgos (2013), Poncela and García Ferrer (2014), Bermingham and D'Agostino (2014), and Stock and Watson (2015)).

There are two broad possibilities for forecasting an aggregate; directly and indirectly. Traditionally, direct methods were limited to forecast the aggregate with a scalar model. Within the indirect group, there are two traditional approaches; forecasting the components with scalar models and then aggregate the forecasts, or use a multivariate model to obtain the components' forecasts and then aggregate the results.

Theoretical results indicate that when the data generation process (DGP) is known, the multivariate approach is at least as accurate as the other two (see Kohn (1982) and Lütkepohl (1987)). As showed by Kohn (1982) and Clark (2000), the direct approach will also be efficient under some very special conditions (conditions for the efficiency of the direct forecast, CEDF) that generally are not satisfied by the data. When these conditions do not apply, direct forecasts introduce invalid restrictions on the DGP, affecting the forecasting accuracy of the aggregate.

Though the multivariate approach would be theoretically preferred, when the DGP is unknown and the sample size finite, the forecast error depends on the number of parameters estimated (see Lütkepohl (1987) and Yamamoto (1981) for the details). Hence, a trade-off between parameter estimation uncertainty and efficiency gains derived from specifying the disaggregated system appears in applications with real data. In particular, when the number of components is large, the advantages of the multivariate approach may be completely lost due to the effects of estimation uncertainty on the forecasts accuracy. Furthermore, this approach is not even feasible for a sufficiently large number of components.

In this thesis, since we are interested in modeling and forecasting all the disaggregates (not

only the aggregate), a great amount of estimation effort is inevitable. Giacomini and Granger (2004) propose a way for dealing with the informational losses vs. estimation uncertainty trade-off. Basically, they show that using a restricted multivariate model for all the disaggregates — a model that does not consider all the series in every equation but only the relevant ones — may solve the problem. In their case of interest (spatially correlated variables) the restrictions are known beforehand. When restrictions are unknown, they may be determined by some statistical procedure, at the cost of introducing an additional source of uncertainty in the model. Depending on the procedure's ability to detect the restrictions, it may be the case that the restricted multivariate approach still outperforms other possible alternatives.

Hendry and Hubrich (2011) stress the relevance of the informational losses vs. estimation uncertainty trade-off by showing that estimation uncertainty is one of the main causes of the difference in the forecast errors between direct and indirect methods. They develop a methodology for imposing the restrictions when they are unknown. The cornerstone of their proposal is that including disaggregated information in the model for the aggregate cannot lessen, but may improve, the predictability of the aggregate (see, also, Hendry and Hubrich (2005)). Their proposal is to include some (selected) components in a single-equation model for the aggregate — i.e; to make use of disaggregated information instead of disaggregated forecasts.

Another alternative for dealing with informational-loses vs. the estimation uncertainty trade-off is the consideration of common features, as proposed by Espasa and Mayo-Burgos (2013). Their suggestion is to try to discover blocks of components that share unique common features (trends and cycles), and then to include the restrictions implied by those commonalities in single-equation models for the components. The search for those blocks is carried out by performing common features tests between all the $N(N-1)/2$ pairs that exist in a set of $N$ components. As the authors argue, the level of disaggregation used in this procedure must be the maximum available since *ad-hoc* sub-aggregates may add up series that do not share common features. Castle and Hendry (2010) also highlight the importance of including long and short-run common features restrictions in the individual models for the components, as proposed by Espasa and Mayo-Burgos (2013).

The proposal of Espasa and Mayo-Burgos (2013) may have three advantages with respect to the component selection proposed by Hendry and Hubrich (2011). First, restrictions implied by the existence of common features are explicitly tested, and from these tests structural economic

interpretations may be derived, putting the economic theory closer to the forecasting procedure. In addition, the pairwise approach takes into account possible cointegration and common cycles relationships which, as shown by Clark (2000) and Vahid and Issler (2002), respectively, should not be neglected in forecasting exercises. Third, this procedure allows modeling and forecasting all the disaggregates (not only the aggregate), generating valuable information for economic and business decision making. Finally, Espasa and Mayo-Burgos's strategy does not exclude the component selection approach of Hendry and Hubrich (2011). After specifying the common features, model selection can be used to see which common features and disaggregates are relevant in each component's equation.

The problem of how to impose unknown restrictions in multivariate models is also present in the Dynamic Factors Models (DFM) literature. Boivin and Ng (2006) find that if the data contain non-pervasive factors (factors that are common only to a reduced subset of series), the choice of the data from which the factors are extracted is not innocuous. Results are more accurate when factors are extracted from data that is informative about them. Along the same lines, Espasa and Mayo-Burgos (2013) find that the forecasting performance of DFM is improved if the factor is estimated from series that have it. As a third example, Beck et al. (2015), working in the context of price setting, find that if the loading matrix has a block structure, factors are more precisely estimated if this structure is considered in the estimation process. Finally, Castle and Hendry (2010) also acknowledge the importance of estimating common features from informative data.

Boivin and Ng (2006) propose an ad-hoc procedure that consists of pre-grouping the series in broad categories and extracting the factors from those groups. This procedure is also adopted by Clark (2006). Results in Espasa and Mayo-Burgos (2013) suggest that this pre-grouping strategy will not work well for disaggregated prices because common features are not distributed by 'broad categories'. Beck et al. (2015) do not need to estimate the block structure because they assume they know it, but this may not be a realistic assumption in many cases.

Note that the issues about non-pervasive common features are not only of theoretical interest; a proper estimation of the factors may change dramatically the conclusions about the estimated dynamic properties of aggregate and sectoral prices, and therefore the implications for economic policy and business decisions. In fact, from their empirical application to European inflation, Beck et al. (2015) find that previous conclusions about the properties of sectoral components

should be modified. Specifically, they find the sectoral components to be less volatile than previously estimated, to explain much lees variance of inflation series, and to show little persistence. Then, according to their results, sector specific components are not the main driving force of inflation as previous studies suggested. Country and regional factors are the main drivers of inflation.

A central aim of this thesis is to show that the pairwise strategy initially proposed by Espasa and Mayo-Burgos (2013) can be used as an objective method to discover blocks of components sharing single common features. That is, blocks of components can be constructed by looking for common features between all the $N(N-1)/2$ pairs of series and then grouping those that share a unique common feature. Our contribution in that respect consists of providing the statistical properties of the strategy using analytic and Monte Carlo procedures. This is the main contribution of chapter II, in which we focus just on common trends.

We focus in the case of fixed $N$ and $T$ going to infinity. The asymptotic validity of the pairwise strategy only requires $T$ going to infinity. In our view, this is strength for we do not need to assume that the number of components of an aggregate goes to infinity in order to achieve consistency. Such an assumption could be considered as quite unrealistic. We argue, however, that our theory can be extended for the case of both $T$ and $N$ going to infinity. An interesting result in this respect is that we do not need the common features to be pervasive. That is, the size of the single-feature subsets can grow at a lower rate than $N$.

The other contributions of that chapter are related to the specification of the bi-variate models for the $N(N-1)/2$ cointegration tests, the procedure for building the single-equation models for the basic components, and the strategy for discovering subsets of basic components with a single common trend.

In chapter III, we extend the procedure to make it robust when applied to real data which may be subject to irregularities and/or short samples problems. The robustification concerns both, the construction of the 'single-trend' subsets and the estimation of the single equation models. Additionally, we extend the search of cointegration by the inclusion of weakly exogenous variables and the admission of more general short run dynamics as suggested by Aron and Muellbauer (2013). In chapter III, we also add some 'quality' controls to the estimated cointegration relationships to be accepted as 'good' ones, and illustrate the procedure applying it to the US CPI.

In chapter IV we extend the pairwise procedure with the consideration of common cyclical features. The contributions of this chapter with respect to Espasa and Mayo-Burgos (2013) are several. First, we analytically justify the validity of proceeding in a pairwise fashion for discovering 'single-cycle' subsets. Second, we use a more general definition of common cycles and admit that the 'single-cycle' subsets can be overlapped with the 'single-trend' ones. Third, by means of a number of Monte Carlo exercises, we study the properties of the pairwise approach applied to common cycles and compare its behavior with a DFM alternative. Finally, we do not restrict the search to the largest subset but look for all possible subsets of series sharing a single cycle. Additionally, in the short chapter V, we consider three relevant issues for empirical applications of the pairwise approach with common cycles: the inclusion of a strongly exogenous variable, the extraction of the common cycle inside the 'single-cycle' subsets, and the use of the general short run dynamics suggested by Aron and Muellbauer (2013).

In chapter VI we complete the application of the pairwise procedure to the US CPI initiated in chapter III and perform a new application to the US IPI. Finally, chapter VII is devoted to the general conclusions. In this final chapter we summarize the conclusions of previous ones and analyze the limitations and possible future research lines that derive from this thesis.

# Chapter II

# Discovering common trends in a large set of disaggregates: statistical procedures and their properties

## II.1 Introduction

As argued in the introduction of this thesis, a possibility for dealing with the informational-loses vs. estimation uncertainty trade-off that arises when dealing with disaggregates is the consideration of common features, as proposed by Espasa and Mayo-Burgos (2013). The authors suggest to try to discover blocks of components that share unique common features (trends and cycles), and then include the restrictions implied by those commonalities in the building process of single-equation models for the components. The search for those blocks is carried out by performing common features tests between all the $N(N-1)/2$ pairs that exist in a set of $N$ components. As they argue, the level of disaggregation used in this procedure must be the maximum available since *ad-hoc* sub-aggregates may add up series that do not share common features. Castle and Hendry (2010) also highlight the importance of including long and short-run common features restrictions in the individual models for the components, as proposed by Espasa and Mayo-Burgos (2013).

The main aim of this chapter — in which the consideration of common features is restricted to common trends — is to show that the pairwise strategy initially proposed by Espasa and Mayo-Burgos (2013) can be used as an objective method to discover blocks of components

sharing single common features. That is, blocks of components can be constructed by looking for common trends between all the $N(N-1)/2$ pairs that exist in a set of $N$ components, and then grouping those that share a unique common trend. Our main contribution consists of providing the statistical properties of the strategy using analytic and Monte Carlo procedures.

We show that the probability of finding cointegration between all possible pairs in a subset of components that share a unique common trend does not depend on the number of series and tends to $(1-\varphi)$ in large samples ($T \Rightarrow \infty$), where $\varphi$ is the nominal size used in the Johansen's trace test (see theorem 1). Monte Carlo experiments confirm this result. In simulation exercises, we also compare the performance of the pairwise approach with an alternative based on Dynamic Factor Models and find that the former performs better in cases in which the proportion of series that share a unique common trend is relatively small, say, 25% of the whole set of series.

The other contributions of this chapter are related to the specification of the bi-variate models for the $N(N-1)/2$ cointegration tests, the building procedure for the single-equation models for the basic components and the strategy for discovering subsets of basic components with a common trend.

Concerning cointegration, we perform bi-variate Johansen's tests, which among other things, avoid the problem in Espasa and Mayo-Burgos (2013) regarding the direction in which one should estimate the Engle-Granger equations. We also analyze the importance of a correct specification of the dynamic structure for different 'types of pairs' and conclude that it may be critical for cointegration results and for the performance of the pairwise strategy.

For sets of series that share a single common trend, we compare the pairwise strategy with a full model approach in which cointegration is tested in a single model that includes all the variables. For relatively short samples, the former has larger power to find the true number of cointegration relationships than the latter. This finding derives from the result in Johansen (1995) and Lütkepohl et al. (2003) about the negative dependence of the Johansen's test's power with respect to the number of common trends in the system.

In relation to the single-equation models for the components, we argue that different normalizations of the same cointegration relationships — by implying different adjustment matrices in the full model for all the components — may lead to a reduction in the number of parameters in the single-equation models. Thus, estimation uncertainty reductions and forecasting accuracy improvements may arise from a simple change in the normalization of the cointegration

relationships, what could be exploited in applied work.

We focus in the case of fixed $N$ and $T$ going to infinity. The asymptotic validity of the pairwise strategy only requires $T$ going to infinity. In our view, this is strength for we do not need to assume that the number of components of an aggregate goes to infinity in order to achieve consistency. Such an assumption could be considered as quite unrealistic. However, in §II.4 we give some preliminary arguments supporting that our theory can be extended for the case of both $T$ and $N$ going to infinity. An interesting result in this respect is that we would not need the common trends to be pervasive. That is, the size of the fully cointegrated subsets could grow at a lower rate than $N$. Additionally, we would be able to deal with cases where $N/T \to \infty$.

An approach that is related to the pairwise procedure is the one proposed by Pesaran (2007), who develops a strategy for testing output and growth convergence across countries. Output convergence between two countries implies that log GDPs are cointegrated with cointegrating vector $[1, -1]$, and without trend in the cointegration relationship. This can be tested by performing unit root tests on the log difference of the two GDPs and checking the significance of the deterministic trend. For a group of $N$ countries, the testing strategy consists of performing unit root tests for all the $N(N-1)/2$ differences between pairs of log GDPs. Pesaran (2007) shows that, under the null of convergence, the fraction of false unit root conclusions tends to $\varphi$ as $N$ and $T$ go to infinity (where $\varphi$ is the nominal size used in the unit root tests).

There is an important difference between Pesaran's (2007) objective and ours. Given a group of $N$ series, Pesaran (2007) is interested in testing the null hypothesis that (almost) all log differences between the series in the data set are stationary without deterministic trend. In contrast, given the same group of series, our objective is to discover subgroups that share a single common trend. In our approach, not (almost) all the series need to share the trend and there may be many subgroups in which all the elements share a unique trend. This is what one could expect when dealing with basic components. In Pesaran's study, this would imply that there would be different convergences by groups of countries. Another important difference is that we do not want to restrict the cointegration vectors to be $[1 - 1]$, so that the use of log differences will not work in our case.

The rest of the chapter is organized as follows. In §II.2, we study the statistical foundations of the pairwise procedure, including a justification for a possible indirect forecasting strategy.

In §II.3, we perform a Monte Carlo experiment to confirm the results of previous section and compare the performance of our procedure with that of an alternative based on Dynamic Factor Models. Finally, in §II.4 we give an preliminary discussion for the case of $N \to \infty$, and§III.7 is devoted to the conclusions.

## II.2 Statistical foundations of the pairwise procedure

The general framework for the models we work with is given by a VAR model for all the $N$ components of an aggregate that are assumed to be $I(1)$:

$$\begin{aligned} X_t = \mu_t + \Pi_1 X_{t-1} + ... + \Pi_k X_{t-k} + \epsilon_t \ \Rightarrow \\ (I_n - \Pi_1 L - ...\Pi_k L^k) X_t = \mu + \epsilon_t \Rightarrow \Pi(L) X_t = \mu + \epsilon_t, \end{aligned} \tag{II.1}$$

where $X_t$ is a $N \times 1$ vector; $\Pi_i$ are $(N \times N)$ matrices of coefficients; $\epsilon_t$ is a Gaussian white noise iid vector; $\mu_t$ contains the deterministic components (constants and trends); $\Pi(z)$ is the characteristic polynomial; and $L$ is the lag operator. If the system is cointegrated, it can be rewritten as a Vector Equilibrium Correction Model (VEqM):

$$\Delta X_t = \mu_t + \alpha\beta' X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1} \Delta X_{t-k-1} + \epsilon_t, \tag{II.2}$$

where $\alpha$ and $\beta$ are $N \times r$ matrices, with $0 < r < N$, $r$ being the number of cointegration relationships, $\alpha\beta' = -I_n + \Pi_1 + ... + \Pi_k$, and $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$.

The data structure for which our procedure is designed can be summarized in three assumptions:

**Assumption 1** *The $N$ components are generated by the VAR eq. (III.5).*

**Assumption 2** *The $N$ components are $I(1)$.*

**Assumption 3** *There is, at least, one subset of $n_1$ components that share a unique common trend.*

**Assumption 4** *The residuals of eq. (III.5) are iid and normally distributed.*

Assumption 2 avoids the complication of dealing with $I(2)$ patterns or seasonal unit roots in cointegration analysis, and rules out the 'trivial' cointegration relationships that will appear

when there are $I(0)$ components. As we argue in chapter III this is a sensible assumption for the components of the US CPI. In chapter VI we argue that the assumption of no seasonal unit roots could be unrealistic for the US Industrial Production Index, what, among other reasons, leaded us to use seasonal adjusted series, see that chapter for a more detailed discussion.

Assumption 3 gives relevance to our objective of discovering 'fully-cointegrated' subsets, and assumption 4 is necessary for the Johansen's maximum likelihood procedure.

**Remark 1** *Instead of assumption 2 we could assume that at least $n_1$ components are $I(1)$ and the others are either $I(1)$ or $I(0)$. This new assumption would require testing the significance of the cointegration relationships' coefficients ($\beta$). This is how we proceed in the empirical applications of chapter III and chapter VI.*

**Remark 2** *Assumption 4 is related to the residuals, not to the first differences of the components. Although this distinction is not relevant for the simulations of this chapter, it is important for the empirical applications of chapter III and chapter VI. In chapter III we design a procedure for dealing with outliers and breaks in the framework of the pairwise approach. By allowing for location shifts in the distribution of the first difference of the components, we do not need to assume normality of $\Delta X_{it}$. Our only requirement is that normality can be achieved after correcting for a few location shifts.*

**Remark 3** *Cheung and Lai (1993) show that the Johansen's trace test is not substantially affected by skewness and/or excess kurtosis of the residuals. Therefore, even if the assumption that after admitting location shifts $\Delta X_{it}$ is normal, is not valid, we do not expect a substantial deterioration of our procedure.*

For a chosen common feature — common trend, common cycle, etc. — the pairwise strategy looks for blocks of components that share just one common feature of the selected type. For the case of common trends, the procedure requires performing Johansen's cointegration tests between all possible pairs of components. For each pair, a bi-variate VAR model has to be estimated and the lag length determined in each case. Then, the procedure requires constructing subsets in which every series is cointegrated with all the others. We call these subsets *Fully Cointegrated* subsets. This strategy relies on the transitivity of cointegration.

Next, a single equation model for each component can be estimated, including as potential regressors all the possibly relevant (if any) cointegration relationships found in previous step, as

well as each component's own lags and lags of other components. The selection of the relevant regressors can be carried out by the model selection algorithm *Autometrics* (see Doornik (2009)). The resulting modeling procedure is something intermediate between the full vector model — which is unfeasible in our context of large $N$— and the univariate estimation of each component. Finally, the single-equation models can be used for forecasting all the components.

### II.2.1 Asymptotic properties ($T \to \infty$) of the pairwise procedure

We now focus on analyzing the cost, if any, of proceeding by pairs instead of specifying the full multivariate model and performing regular Johansen's cointegration tests.

To fix ideas, assume that we are dealing with a macro-variable composed by, say, $N = 100$ basic components. The total number of pairs is $N(N-1)/2 = 4950$, so that we initially need to perform this number of cointegration tests. Assume, further, that a subset of much smaller dimension —$n_1$—, say $n_1 = 40$, within the 100 components share a unique common trend, and the other $N - n_1 = 60$ have their own trends. We will use the notation $n_1$ for both, to indicate the size and to label the 'fully cointegrated subset'.

The ideal properties of the procedure are: (1) Cointegration tests between all possible pairs in $n_1$ (which are $n_1(n_1-1)/2 = 780$) should indicate the existence of a cointegration relationship, and (2) no series outside $n_1$ should be — wrongly — included in $\hat{n}_1$.

Condition 2 seems the most relevant since including wrong series in a fully cointegrated set will bias the forecasts of the components in that set. Additionally, if the proportion of wrong series is large, the procedure will collapse. On the other hand, not fulfilling condition 1 will generate efficiency losses but will not bias the forecasts.

Since we are performing Johansen's tests, the procedure inherits its asymptotic properties. There are, however, two specific features that deserve special attention: multiple testing and estimation of partial models.

#### II.2.1.1 The general problem of multiple testing

In the regular framework in which there is not repeated hypothesis testing, the probability of not false rejecting the null is $1 - \varphi$ (with $\varphi$ being the nominal size of the test). When $m$ tests are performed, assuming that they are independent, the probability of not making any false rejection reduces to $(1-\varphi)^m$, and the probability of making at least one error is $1 - (1-\varphi)^m$, which rapidly increases with $m$.

Several approaches have been proposed for controlling type I error rates in multiple testing frameworks, among which those that try to control either the Family Wise Error Rate ($FWER$) or the False Discovery Rate ($FDR$) seem to be the most popular (see, e.g., Benjamini and Hochberg (1995)). Defining $V$ as the number of true null hypotheses that were — wrongly — rejected, and $R$ as the total number of rejections: $FWER = P(V \geq 1)$, and $FDR = E(V/R)$. When there are some true null hypotheses, any procedure that controls $FWER$ also controls $FDR$; that is, controlling $FWER$ requires tighter adjustments (see Benjamini and Hochberg (1995)).

The Bonferroni correction adjusts all p-values in a single step for ensuring that $FWER \leq \varphi$. The principle is simple; if $m$ tests are performed, each of them has to be rejected whenever its p-value is smaller than $\varphi/m$ (strictly speaking, the corrected p-value is $1-(1-\varphi)^{1/m}$, but this magnitude is often approximated by $\varphi/m$). This ensures that $FWER \leq \varphi$.

The argument for such a correction is the following. Assume that we have performed two tests whose null hypotheses are $H_0^1, H_0^2$, and call $FR_1$ and $FR_2$ the events of falsely rejecting $H_0^1$ and $H_0^2$, respectively. Then, $FWER = P(FR_1 \cup FR_2) = P(FR_1) + P(FR_2) - P(FR_1 \cap FR_2) = 2\varphi - P(FR_1 \cap FR_2)$. If the events are independent, then $P(FR_1 \cap FR_2) = \varphi^2$ and the Bonferroni correction will deliver $FWER = \varphi$. When $P(FR_1 \cap FR_2) > \varphi^2$, $FWER < \varphi$ and Bonferroni corrections will be too stringent, even if we are interested in testing whether, at least, one individual hypothesis is false (this is frequently called the *universal null hypothesis*).

A case of interest may be when $P(FR_1|FR_2) \simeq 1$ (or $P(FR_2|FR_1) \simeq 1$), such that $P(FR_1 \cap FR_2) \simeq \varphi$. In this case, $FWER \simeq \varphi$, and there is no need to adjust p-values, even if the relevant hypothesis is the *universal* one. Using the principle of inclusion and exclusion for probability, this last argument can be generalized for the case of $m$ tests. That is, if the probability of wrongly rejecting any combination of the $m$ hypothesis at the same time is close to $\varphi$, it can be easily seen that $FWER = \varphi$, and there is no need for correcting p-values.

#### II.2.1.2 Multiple testing in the pairwise approach

Since the pairwise procedure involves a large number of cointegration tests (4950 for $N = 100$), it may be thought to raise the probability of false rejection. We analyze this issue for the three different types of pairs: i) Both series belong to $n_1$ so that, in the context of Johansen's tests,

the problem is rejecting the true hypothesis of $r = 1$ (one cointegration relationship) [1]; ii.a) only one series belongs to $n_1$ so that the problem now is rejecting the true hypothesis of $r = 0$; and ii.b) none of the series belongs to $n_1$ so that the problem again consists of rejecting the true hypothesis of $r = 0$.

As we argue below, multiple testing does not occur for pairs in i. For pairs in ii.a, multiple testing may occur, but our procedure includes an automatic control for this issue. Still, even if this automatic control does not work, the possible inflated rejection frequency of the true null hypothesis ($r = 0$) is tolerable for widely general dataset configurations. Finally, for pairs in ii.b, multiple testing occurs, but the *full cointegration* requirement automatically controls this issue, and false null rejections of $r = 0$ are not an issue for this group of series.

*False rejection of* $r = 1$

In the Johansen procedure, the null hypothesis $r = 0$ and $r = 1$ are usually tested sequentially. Since the asymptotic power of Johansen's test is 1, finding no cointegration between pairs in $n_1$ is not an issue in large samples. The problem is, therefore, false rejecting $r = 1$ in favor of $r = 2$. If the tests were independent, the probability of finding one common trend between *all* series in $n_1$ would be $(1 - \varphi)^{n_1(n_1-1)/2}$, which quickly decreases with $n_1$. But, clearly, the tests are not independent. In fact, theorem 1 below indicates that these tests are asymptotically equivalent in the sense that the probability of obtaining the same result in all of them goes to 1 as $T$ goes to infinity.

**Theorem 1** (Asymptotic equivalence of pairwise cointegration tests in a fully cointegrated set). *Given a set of $Q$ $I(1)$ pairwise cointegrated series (i.e., there are $Q - 1$ cointegration relationships among them and a single common trend), the probability of obtaining the same result in all the $Q(Q-1)/2$ pairwise Johansen's trace tests tends to 1 in as $T \to \infty$.*

**Proof** *The first step in Johansen's procedure is to concentrate the model with respect to $\alpha\beta'$, what is done by regressing $\Delta Y_t$ and $Y_{t-1}$ on $(\Delta Y_{t-1}, ..., \Delta Y_{t-k+1})$. These auxiliary regressions give the residuals $R_{0t}$ and $R_{1t}$, respectively, and the matrices $S_{ij}$ are defined as $T^{-1}R_iR_j'$, where $R_i$ is a $n \times T$ matrix. For $n = 2$, the likelihood ratio test for the null $r = 1$ vs. $r = 2$ is: $-Tln(1 - \hat{\lambda}_2)$, where $\hat{\lambda}_2$ is the smallest eigenvalue of the generalized eigenvalue problem:*

$$(S_{10}S_{00}^{-1}S_{01})v = \lambda S_{11}v, \tag{II.3}$$

[1] Although Johansen's test is sequential, the asymptotic probability of not rejecting $r = 0$ is asymptotically 0, for asymptotic power is 1.

*whose eigenvalues are the solution of,* $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$.

*Let* $\boldsymbol{X}_t$ *be the vector containing the series in* $n_1$. *From the Granger Representation Theorem, disregarding deterministic terms, the cointegrated VAR can be written as:*

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + C(1)\sum_{i=1}^{t}\epsilon_t + C^*(L)\epsilon_t,$$

*where* $C(1) = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$, *has rank* $n_1 - r$, *and* $C^*(L)$ *is a stationary lag polynomial matrix. Under full cointegration,* $r = n_1 - 1$ *and the rank of* $C(1)$ *is 1. Therefore, individual series in* $n_1$ *can be written as:*

$$X_{mt} = \delta_m CT_t + w_{mt}; \quad m = 1, ..., n_1, \tag{II.4}$$

*where* $CT_t$ *is a* $I(1)$ *process and* $w_{mt}$ *are stationary ones. From eq. (II.4), any series in* $n_1$ *can be expressed as* $X_{mt} = \gamma_{mq}X_{qt} + \eta_{mr,t}$, *with* $\eta_{mt} \sim I(0)$, *and* $X_{qt}$ *being other series in* $n_1$.

*Assume that the aforementioned matrices* $S_{ij}$ *refer to the vector ,* $Y_t = [X_{1t}, X_{2t}]'$. *We now derive the test statistic for any other pair in* $n_1$ *given the one for* $Y_t$. *Let* $Y_t^* = [X_{it}, X_{jt}]'$, *and write:*

$$\begin{aligned} X_{it} &= \gamma_{i1}X_{1t} + \eta_{i1,t}, \\ X_{jt} &= \gamma_{j2}X_{2t} + \eta_{j2,t}, \end{aligned} \tag{II.5}$$

*The auxiliary regressions are now:*

$$\begin{aligned} \Delta Y_t^* &= B_0^*\Delta\tilde{Y}_{t-1}^* + R_{0t}^* \\ Y_{t-1}^* &= B_1^*\Delta\tilde{Y}_{t-1}^* + R_{1t}^*, \end{aligned} \tag{II.6}$$

*where* $\Delta\tilde{Y}_{t-1}^* = [\Delta Y_{t-1}^{*'}, \Delta Y_{t-2}^{*'}, ..., \Delta Y_{t-p^*}^{*'}]$, *and* $B_i^*$ *is a* $2 \times p^*$ *matrix of coefficients. Note now that eq. (II.6) can be written as:*

$$\begin{aligned} \Gamma\Delta Y_t + \Delta\eta_t &= B_0^*(\Gamma\Delta\tilde{Y}_{t-1} + \Delta\tilde{\eta}_{t-1}) + R_{0t}^* \\ \Gamma Y_{t-1} + \eta_{t-1} &= B_1^*(\Gamma\Delta\tilde{Y}_{t-1} + \Delta\tilde{\eta}_{t-1}) + R_{1t}^*, \end{aligned}, \tag{II.7}$$

*where* $\Gamma = \begin{bmatrix} \gamma_{i1} & 0 \\ 0 & \gamma_{j2} \end{bmatrix}$, *and* $\Delta\tilde{\eta}_{t-1} = [\Delta\eta_{t-1}', \Delta\eta_{t-2}', ..., \Delta\eta_{t-p^*}']$.

*Let* $B_i^* = \Gamma B_i\Gamma^{-1} + b_i$, *with* $b_i$ *an adequate* $2 \times p^*$ *matrix. To save notation, we assume*

$p^* = p$, but the argument only requires $p^* \geq p$. Equations in eq. (II.7) become:

$$\begin{aligned} \Gamma\Delta Y_t + \Delta\eta_t &= \Gamma B_0 \Delta\tilde{Y}_{t-1} + b_0\Gamma\Delta\tilde{Y}_{t-1} + B_0^*\Delta\tilde{\eta}_{t-1} + R_{0t}^* \\ \Gamma Y_{t-1} + \eta_t &= \Gamma B_1 \Delta\tilde{Y}_{t-1} + b_1\Gamma\Delta\tilde{Y}_{t-1} + B_1^*\Delta\tilde{\eta}_{t-1} + R_{1t}^* \end{aligned} \tag{II.8}$$

Recall that the original $R_{it}$ are $R_{0t} = \Delta Y_t - B_0\Delta\tilde{Y}_{t-1}$ and $R_{1t} = Y_{t-1} - B_1\Delta\tilde{Y}_{t-1}$ . Then, solving eq. (II.8) for $R_{it}^*$:

$$\begin{aligned} R_{0t}^* &= \Gamma R_{0t} + \epsilon_{0t} \\ R_{1t}^* &= \Gamma R_{1t} + \epsilon_{1t} \end{aligned},$$

where, $\epsilon_{0t}$ and $\epsilon_{1t}$ are the following stationary processes:

$$\begin{aligned} e_{0t} &= -[b_0\Gamma\Delta\tilde{Y}_{t-1} + B_0^*\Delta\tilde{\eta}_{t-1} - \Delta\eta_t] \\ e_{1t} &= -[b_1\Gamma\Delta\tilde{Y}_{t-1} + B_1^*\Delta\tilde{\eta}_{t-1} - \eta_{t-1}] \end{aligned}.$$

Hence, the new $(2 \times 2)$ matrices $S_{ij}^*$ are:

$$S_{ij}^* = T^{-1}(\Gamma R_i + \epsilon_i)(\Gamma R_j + \epsilon_j)',$$

then;

$$S_{11}^* = T^{-1}[\Gamma R_1 R_1'\Gamma' + \Gamma R_1\epsilon_1' + \epsilon_1 R_1'\Gamma' + \epsilon_1\epsilon_1'] \tag{II.9}$$

$$S_{11}^* = T^{-1}[\Gamma R_1 R_1'\Gamma' + \Gamma R_1\epsilon_1' + \epsilon_1 R_1'\Gamma' + \epsilon_1\epsilon_1'] \tag{II.10}$$

In eq. (II.10), all terms inside the brackets are $Op(T)$ except for $\Gamma R_1 R_1'\Gamma'$, which is $Op(T^2)$. Thus, $S_{11}^*$ is $Op(T)$ and its long-run behavior is dominated by $\Gamma R_1 R_1'\Gamma'$. That is, $S_{11}^* \to \Gamma R_1 R_1'\Gamma'$ as $T \to \infty$. The remaining $S_{ij}^*$ are $Op(1)$ and can be written as

$$S_{ij}^* = \Gamma S_{ij}\Gamma + \Omega_{ij}, \text{ for } (i,j) \neq (1,1), \tag{II.11}$$

where $\Omega_{ij} = T^{-1}[\Gamma R_i\epsilon_j' + \epsilon_i R_j'\Gamma' + \epsilon_i\epsilon_j']$ is $Op(1)$ for $(i,j) \neq (1,1)$.

The new eigenvalue problem is: $(S_{10}^* S_{00}^{*-1} S_{01}^*)v^* = \lambda^* S_{11}^* v^*$. Using eq. (II.10) and eq. (II.11),

*we get:*

$$[(\Gamma S_{10}\Gamma' + \Omega_{10})(\Gamma S_{00}\Gamma' + \Omega_{00})^{-1}(\Gamma S_{01}\Gamma' + \Omega_{01})]v^* = \lambda^*(\Gamma S_{11}\Gamma')v^*. \tag{II.12}$$

*Note that* $(\Gamma S_{00}\Gamma' + \Omega_{00})^{-1}$ *can be written as;*

$$(\Gamma S_{00}\Gamma' + \Omega_{00})^{-1} = c(\Gamma S_{00}\Gamma')^{-1} + \tilde{\Omega}_{00} = c\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \tilde{\Omega}_{00} \tag{II.13}$$

*where* $c = \dfrac{|\Gamma S_{00}\Gamma'|}{|\Gamma S_{00}\Gamma' + \Omega_{00}|}$, *and* $\tilde{\Omega}_{00} = \dfrac{1}{|\Gamma S_{00}\Gamma' + \Omega_{00}|}Adj[\Omega_{00}]$ *(note that* $0 < c \leq 1$*).*

*Hence, plugging eq. (II.13) into eq. (II.12) we get:*

$$[(\Gamma S_{10}\Gamma' + \Omega_{10})(c\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \tilde{\Omega}_{00})(\Gamma S_{01}\Gamma' + \Omega_{01})]v^* = \lambda^*(\Gamma S_{11}\Gamma')v^*,$$

*from where:*

$$[c\Gamma S_{10}S_{00}^{-1}S_{01}\Gamma' + (\Gamma S_{10}\Gamma'\tilde{\Omega}_{00} + c\Omega_{10}\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \Omega_{10}\tilde{\Omega}_{00})(\Gamma S_{01}\Gamma' + \Omega_{01})]v^* = \lambda^*(\Gamma S_{11}\Gamma')v^*,$$

*and,*

$$[c\Gamma S_{10}S_{00}^{-1}S_{01}\Gamma' + \Psi]v^* = \lambda^*(\Gamma S_{11}\Gamma')v^*, \tag{II.14}$$

*where,* $\Psi$ *is* $Op(1)$ *and its expression is:* $\Psi = c\Gamma S_{10}S_{00}^{-1}\Gamma^{-1}\Omega_{01} + (\Gamma S_{10}\Gamma'\tilde{\Omega}_{00} + c\Omega_{10}\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \Omega_{10}\tilde{\Omega}_{00})(\Gamma S_{01}\Gamma' + \Omega_{01})$.

*Left multiplying eq. (II.14) by* $\Gamma^{-1}$ *we obtain:* $[cS_{10}S_{00}^{-1}S_{01}\Gamma' + \Gamma^{-1}\Psi]v^* = \lambda^*(S_{11}\Gamma')v^*$. *Now, let* $\underset{\sim}{\Psi} = \Gamma^{-1}\Psi\Gamma^{-1}$, *to get:*

$$[cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi}]\Gamma'v^* = \lambda^* S_{11}\Gamma'v^*. \tag{II.15}$$

*Comparing eq. (II.15) with eq. (II.3), we can make three considerations:*

*i) If* $X_i \equiv X_1$ *and* $X_j \equiv X_2$, *we get* $\Psi = 0$, $c = 1$ *and* $\Gamma = I$, *so we recover the original problem.*

*ii) In the extremely unlikely case that* $\Omega_{ij} = 0$ *— for* $(i,j) \neq (1,1)$ *—, we get* $\Psi = 0$ *and* $c = 1$, *so that the eigenvalue problem would be:* $[S_{10}S_{00}^{-1}S_{01}\Gamma']v^* = \lambda^*(S_{11}\Gamma')v^*$, *the solution of which is* $\lambda^* = \lambda$ *and* $v^* = \Gamma'v$. *Hence, even in small samples, the cointegration test statistic*

*would be exactly the same as the one for the pair* $(X_1, X_2)$.

*iii) In the general case that* $\Omega_{ij} \neq 0$, *we will have* $\Psi \neq 0$ *and* $c \neq 1$. *Note that the eigenvalues of the problem eq. (II.15) are the solutions of the second-order polynomial in* $\lambda^*$ $|\lambda^* S_{11} - (cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi})| = 0$.

*Focus on the general case that* $\Omega_{ij} \neq 0$. *As Johansen (1995) shows, the test statistic* $-T\sum_{r+1}^{p} ln(1-\lambda_i)$ *converges to a non-standard distribution that does not depend on* $S_{00}$. *Given that* $S_{11}$ *is* $Op(T)$ *and the other matrices are* $Op(1)$, *the asymptotic behavior of* $\lambda$ *and* $\lambda^*$ *is dominated by the same terms. To see this, let* $\Theta = S_{10}S_{00}^{-1}S_{01}$, *and* $\Theta^* = cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi}$. *The original eigenvalues* $\lambda_1$ *and* $\lambda_2$ $(\lambda_1 > \lambda_2)$*are the roots of the polynomial:*

$$\lambda^2|S_{11}| + \lambda \underbrace{(s_{12}\theta_{21} + s_{21}\theta_{12} - s_{11}\theta_{22} - s_{22}\theta_{11})}_{B} + \underbrace{(\theta_{11}\theta_{22} - \theta_{21}\theta_{12})}_{C} = 0, \tag{II.16}$$

*where* $s_{ij}$ *and* $\theta_{ij}$ *are the elements of the matrices* $S_{11}$ *and* $\Theta$, *respectively.*

*Since* $B < 0$,

$$\lambda_2 = \frac{-B - \sqrt{B^2 - 4|S_{11}|C}}{2|S_{11}|} = \frac{G}{2|S_{11}|}. \tag{II.17}$$

*If the series are cointegrated* $|S_{11}| \sim Op(T)$, *and since* $B \sim Op(T)$, *the expression under the square root is dominated by* $B^2$, *and* $G \to 0$.

*Now, replace* $\theta_{ij}$ *by* $\theta_{ij}^*$ *in eq. (II.16) to get* $B^*$, $C^*$ *and* $G^*$. *Since* $\theta_{ij}$ *and* $\theta_{ij}^*$ *are* $Op(1)$, *the asymptotic behavior of* $G^*$ *is the same as that of* $G$, *for the expression under the square root is also dominated by* $B^{*2}$, *which is determined by the same* $s_{ij}$*'s as* $B$.■

Another way to interpret this theorem could be: *transitivity is a property not only of cointegration, but also of cointegration tests.*

The intuition for this result is that, asymptotically, the $N(N-1)/2$ cointegration tests are tests for one versus no common trend, which can be seen as unit root tests for the estimated common trend. Since this trend is the same for all series, we have $N(N-1)/2$ estimations of the same trend, which tend toward the same process as $T$ goes to infinity.

Implications from theorem 1:

a) Let $WR_{ij}$ be the event in which the null of $r = 1$ is wrongly rejected for the pair $(i, j)$.

Theorem 1 implies that the joint probability for any combination of $WR_{ij}$ (for any $i, j$

belonging to $n_1$) is close to $\varphi$. Therefore, even in the case when the hypothesis of interest is the *universal* one defined in the Bonferroni approach (i.e., false rejecting at least one of the $N(N-1)/2$ hypotheses, which is not necessarily the case of the pairwise procedure), p-values should not be corrected (see table A.1 in appendix A).

b) A procedure for testing the hypothesis that a certain group of series share a single common trend (similar to, but less restrictive than the hypothesis of interest in Pesaran (2007)) could be: i) Test $r = 0$ vs $r > 0$ in all possible pairs using regular critical values — asymptotic power is 1. ii) If all hypothesis are rejected, test $r = 1$ vs $r = 2$ in all pairs, store the maximum test statistic and compare it to the regular critical values.Reject the null of a unique common trend if the maximum statistic is larger than the critical value. Asymptotically, this procedure delivers the correct size.

c) As argued in the introduction, the procedure in b) is not useful for discovering sub-groups of series sharing a unique common trend. To do this, we group series that are pairwise cointegrated ($r = 1$) using regular critical values. As table A.1 of appendix A shows, the probability of including *all* the correct series is $1 - \varphi$, with $\varphi$ being the nominal size of the individual tests, even when the number of pairwise cointegrated pairs is large.

*False rejection of $r = 0$ when only one of the series belongs to the fully cointegrated set.*

Under assumption 2, the true number of cointegration relationships between one series inside and one series outside a fully cointegrated subset is $r = 0$.

Two comments are relevant for this case. First, since to include a series in $\hat{n}_1$ we require all cointegration tests to find $r = 1$, it is evident that the *universal* null — relevant for Bonferroni corrections — is of no interest at all. What is relevant for the pairwise procedure is the probability of wrongly rejecting all $r = 0$ hypotheses, which, in any case, will be smaller than or equal to $\varphi$.

Second, let $X_{out}$ be a series outside $n_1$ and let $WR_i$ be the event of wrongly rejecting $r = 0$ with the $i^{th}$ series in $\hat{n}_1$. Since for wrongly including $X_{out}$ in $\hat{n}_1$ we need to wrongly reject $n_1$ hypotheses, the probability of including it is $P(WR_1 \cap ...WR_{n_1})$. This probability can be factorized as:

$$P(WR_1 \cap ...WR_{n_1}) = P(WR_1|WR_2, ..., WR_{n_1}) \times ... \; P(WR_{n_1-1}|WR_{n_1}) \times P(WR_{n_1}), \quad \text{(II.18)}$$

where $P(WR_{n_1}) = \varphi$ is the nominal size of the individual tests. Using the extreme assumption that all the conditional probabilities in eq. (II.18) are equal to 1, the probability of wrongly including $X_{out}$ in $\hat{n}_1$ would be $\varphi$, and the expected number of wrong series in $\hat{n}_1$, $E[W]$, would be $(N-n_1)\varphi$. For $N-n_1 = 100$ and $\varphi = 0.01$, $E[W] = 1$. It is in this sense ($E[W]$ is small, even under an extreme and very unlikely assumption) that we tolerate the increased false rejection probability for this type of pairs.

As this was a 'worst case' analysis, the result is just an upper bound for the expected number of wrong series. Interestingly, simulation results show that the actual figures far below this upper bound, indicating that the assumption that all conditional probabilities are equal to one is quite extreme (see §II.3).

*False rejection of $r = 0$ when none of the series belongs to the fully cointegrated set.*

From Johansen's test properties, the probability of finding one cointegration relationship between two non-cointegrated I(1) series — assumption 2 — tends to $\varphi$ as $T \to \infty$. Assuming that tests are independent, we can think of the set of series outside $n_1$ as a random graph with edge probability equal to $\varphi$. Then, a lower bound for the expected number of estimated fully cointegrated sets composed by $K$ series of the $N-n_1$ would be $E[K_{worng}] = C_K^{(N-n_1)}\varphi^{K(K-1)/2}$, which is almost zero for, say, $\varphi = 0.01$, $K > 3$ and moderately large $(N-n_1)$ (see Bollobás and Erdös (1976) for a discussion of *cliques*[2] in random graphs).

Additionally, Matula (1976) shows that the size of the *maximal clique*[3] in a random graph with edge probability $\varphi$ has a strong peak around $2log(N)/log(1/\varphi)$, which is 2 for $N = 100$ and $\varphi = 0.01$. Thus, selecting a low $\varphi$ and disregarding estimated fully cointegrated sets with fewer than three elements constitute a *strong protection* against finding fully cointegrated sets among these series.

Note that $E[K_{worng}]$ is a lower bound since we assume that tests are independent. Notably, simulation results show that the actual number of wrong fully cointegrated sets is close to this bound, meaning that, even though we are performing tests between all the pairs of a set of series, the independence assumption is sensible for this type of pairs (see appendix A.2 and §II.3).

[2]Cliques are sub-graphs in which all nodes are pairwise connected.
[3]The maximal sub-graph in which all nodes are pairwise connected.

#### II.2.1.3 Partial systems

The strategy of performing cointegration tests between all possible pairs of series is justified by the fact that in a set of $n_1$ series that share a unique common trend, there are $n_1 - 1$ cointegration relationships and the series are pairwise cointegrated. Note that this strategy requires partial systems' estimation in the sense that we assume the existence of a full VAR model for all the components but estimate several partial bi-variate systems.

Johansen (1992) states the conditions under which inference for a cointegrated VAR model can be conducted from partial models; if the partial models exclude only weakly exogenous variables for $\beta$ — but condition on them — it is a valid device to make inference for the full system (see, also, Hendry (1995)). Building on this result, Harbo et al. (1998) show how to modify the regular Johansen's test in order to make inference from the partial model when weak exogenity holds. Basically, an additional regressor (the difference of the excluded vector) has to be added to the auxiliary regression for the unexcluded vector, and a new set of asymptotic tables have to be used.

The models considered in the pairwise procedure are partial in the sense that we consider only a subset of variables, but not in the sense of Johansen (1992) and Harbo et al. (1998). That is, we are not seeking to estimate all the cointegration parameters from a bi-variate model (which is impossible). On the contrary, under full cointegration, since every pair of variables is cointegrated, the bi-variate VAR models are complete because all relevant variables are considered endogenous.

In general, the lag length in the bi-variate models will be larger than or equal to that of the full model. The lag length of the partial models will depend on the series we are considering. There are three different cases to distinguish; i) both series have the common trend of $n_1$; ii) only one of them has it; or iii) none of them has the common trend (for a discussion on linear transformations of VAR processes, see Lütkepohl (1984)). Thus, the dynamic structure has to be selected for each pair using some information criteria. Interestingly, when this is done, the power of the pairwise procedure for finding the true number of cointegration relationships $(n_1 - 1)$ is improved with respect to the traditional Johansen's trace test. In §II.3, we show this by means of a small simulation experiment.

### II.2.2 The problem of normalization in forecasting

The cointegrated VAR in eq. (III.5) remains unchanged if we change the original matrices $\alpha$ and $\beta'$ by $\tilde{\alpha} = \alpha H^{-1}$ and $\tilde{\beta}' = H\beta'$, with $H$ being a $r \times r$ normalizing matrix. Although these changes will not have any consequences for the system's dynamic properties, they may change and eventually ease its interpretation.

Two normalizations of interest in a system with a unique common trend between a subset ($n_1$) of the disaggregates of a macro variable may be to express the $n_1 - 1$ cointegration relationships either as deviations with respect to one selected series — normalization *(a)* — or as deviations from the sub-aggregate formed by the series in $n_1$ — normalization *(b)*.

Assume that we choose normalization $a$ and get a very 'simple' matrix $\alpha$ in which each series reacts only to its 'own' cointegration relationship and the normalizing variable is weakly exogenous (i.e., the matrix $\alpha$ is full of zeros except for $r$ coefficients). It can be shown that re-normalizing this system with option $b$ leads to a matrix $\alpha$ with no zero restrictions.

This last observation makes the discussion about normalization relevant to the forecasting stage of the pairwise procedure. The fact that a system with a 'complex' matrix $\alpha$ may be transformed into another with a 'simpler' matrix implies a reduction in the number of variables of the individual forecasting equations without cost in terms of informational losses[4]. We may normalize the system in these two alternative ways and select the cointegration relationships that are relevant to each individual equation. Proceeding in this way, we may get a 'free' reduction in estimation uncertainty, which would translate into an improvement in the forecasting accuracy. Clearly, this is not necessarily the case, but it could be.

Therefore, for each forecasting equation, we may proceed as follows: i) From all the cointegration relationships with the sub-aggregate, select the significant ones. ii) From all the cointegration relationships with a single component, select the significant ones. iii) Keep the simpler model.

This strategy allows us to capture all the long-run information in the system, with the minimum number of parameters, which cannot reduce the forecasting accuracy and may improve it.

Note that this strategy will work properly regardless of the variable we select to normalize the

[4]The change in the normalization can also change the number of elements in matrix $\beta$, but since their estimators are super-consistent, estimation uncertainty is of minor importance in moderate samples.

cointegration matrix. Assume, for example, that the simplest possible specification is a model with normalization *(a)* in which each variable reacts to its 'own' cointegration relationship with respect to, say, $X_{1t}$. Assume further that we choose other series (say, $X_{2t}$) to normalize the cointegration relationships. It can be shown that the only effect of this change is to add just one cointegration relationship to each forecasting equation. This suggests that the effort of normalizing in all possible components to find the one that implies fewer terms in the forecasting equations does not seems to be worth, given the small cost of choosing a random one.

The procedure proposed by Espasa and Mayo-Burgos (2013) for the forecasting stage is to include in each individual equation the cointegration relationship of the corresponding component with the sub-aggregate formed by the series in $n_1$, and no other long-run information. This simple strategy will be correct only in the special case in which applying normalization $b$ to the true DGP, only 'the own' cointegration relationship is relevant for each variable. In more general situations, this approach will omit relevant long-run information.

In summary, when using single equation models for forecasting individual series that have cointegration relationships, different normalization of the cointegration relationships could lead to different forecasting accuracy. Clearly, this is not the case when forecasting by means of a multiequational system.

## II.3 Simulation results for the Pairwise strategy

In this section, we perform two different Monte Carlo experiments. The first one (in §II.3.2) is designed to compare the power properties of the pairwise approach to find the true number of cointegration relationships with respect to the traditional Johansen's trace test (see the discussion in §II.2.1.3 about partial systems). The second experiment is designed with two objectives: assess the analytic results presented in §II.2.1.2, and compare the performance of the pairwise approach with an alternative based on Dynamic Factor Models. These two objectives are covered in §II.3.3, and §II.3.4, respectively. It is important to note that our objective is not making a general comparison between our approach and DFM; we do not want to extract general results. Our goal is much simpler, we just want to evaluate if the usual and simple strategy of estimating factor models by principal components can be used in our framework of interest, namely, relatively large $N$ and relatively small $n_1$.

Before presenting the simulation results, in §II.3.1, we describe the general design of the

simulations, which is similar to both experiments.

### II.3.1 General design of the experiments

We consider two classes of DGP: VEqM and Dynamic Factor Models.

#### DGP 1 - VEqM

The general expression for the VEqCM model of the $N$ series — which can be seen as the whole set of components of a macro variable — is eq. (III.5) with only one lag and $\mu_t = 0$.

We want to simulate a situation in which a subset $n_1$ of the $N$ components share a unique common trend ($n_1 - 1$ cointegration relationships among them), and the rest of the components have their own trends.

Without loss of generality, we set matrix $\beta$ such that:

$$\beta' = \begin{pmatrix} \beta_2 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \beta_3 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \beta_4 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & & \\ \beta_N & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}_{r \times N}$$

This normalization was suggested by Clements and Hendry (1995). Different normalizations change the exact shocks that drive the long-run behavior of the $n_1$ variables, but not the fact that they are determined by $N - r$ shocks and $r$ adjusting mechanisms.

For the sake of simplicity, matrix $\alpha$ is set to have the following structure:

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r} = \begin{pmatrix} \alpha^U \\ \alpha^D \end{pmatrix},$$

where sub-matrix $\alpha^U$ is $n_1 \times r$; sub matrix $\alpha^D$ is a matrix of zeros with dimensions $(N - n_1) \times r$; and the values $\alpha_i$ are taken from the uniform distribution with parameters $[0.15, 0.3]$. These parameters are motivated by results in Espasa and Mayo-Burgos (2013) for CPI series.

In the specification of a DGP, given a selection of $\beta$, the choice of $\alpha$ does affect the properties of the process. We are assuming that the common trend among series in $n_1$ is driven by a single shock, so that the first variable is exogenous and each cointegrating relation affects only one of the remaining variables.

With these structures for the long-run matrices, and setting the $r \times 1$ vector $[\beta_2, ..., \beta_N]'$ equal to $[-1, ..., -1]'$, the series in $n_1$ follow a process similar to that specified by Banerjee and Marcellino (2009) in their *DGP1*, but with two differences[5]. First, we do not assume that the $\alpha_i$s are equal, and second, we add some short-run structure ($\Phi \Delta X_{t-1}$). The objective of including the lagged difference in the model is to avoid the presence of common cycles. Given that $\alpha$ has reduced rank, there exists a matrix $\alpha'_\perp$ such that $\alpha'_\perp \alpha = 0$, so that not including the term $\Phi \Delta X_{t-1}$ would lead to $\alpha'_\perp \Delta X_t = \alpha'_\perp \epsilon_t$, which implies a common cycle structure (the fact that cointegrated VAR(1) processes always present a common cycle structure was first highlighted by Vahid and Engle (1993)). Augmenting the order of the VAR by including the full rank matrix $\Phi$ avoids this issue. For simplicity, it is a diagonal matrix whose diagonal elements are drawn from the uniform distribution with parameters [0.5, 0.8].

At first glance, the selected structure for matrix $\alpha^U$ could seem too simple to be realistic. Note, however, that the complexity of the system cannot be judged from matrix $\alpha$ alone. For example, it can be shown that *DGP 1* is exactly equivalent to a DGP with a 'complete' matrix $\alpha^U$ in which cointegration relationships are normalized with respect to the sub-aggregate formed by the first $n_1$ series. Thus, with such a normalization we would have the same system but it would not be subject to the critique that $\alpha$ is too simple.

### DGP 2 - DFM

The second DGP is a Dynamic Factor Model:

$$\begin{aligned} Y_t &= \Lambda F_t + \epsilon_t, \\ F_t &= F_{t-1} + \eta_t, \end{aligned} \tag{II.19}$$

[5]The authors propose a generalization of the Factor Augmented VAR models to consider cointegration and call the new model FECM. They use this DGP to show an analytical example of the model's properties and to obtain Monte Carlo results.

where $\Lambda$ is a $n \times 1$ vector of factor loadings whose first $n_1$ elements are taken from the uniform distribution with parameters $[0.1, 0.8]$, and the others are 0. $\epsilon_t$ is an $n \times 1$ vector whose first $n_1$ elements are *iid* $N(0, \Sigma)$ with $\sigma_{ii}^2 = (1 - \lambda_i^2)/2$, and the others $N - n_1$ are random walks with no drift such that their first differences are also *iid* $N(0, I)$ . Finally, $\eta_t$ is a *iid* $N(0, 1)$ processes independent of $\epsilon_t$. This structure implies that the first difference of each series has unit variance, and the signal to noise ratio is $\lambda_i^2/(1 - \lambda_i^2)$.

### II.3.2 Power comparisons: Pairwise vs. Johansen's trace test

In this section, we compare the power of the pairwise procedure for finding the true number of cointegration relationships $(n_1 - 1)$ with that of the traditional full system Johansen's trace test (see discussion in §II.2.1.3 about partial systems). For the *DGP 1* described above, we consider the following possibilities for $[N, n_1]$: i) $[6, 2]$. There are six variables and one cointegration relationship between two of them. ii) $[6, 3]$. There are six variables and two cointegration relationships between three of them. iii) $[9, 2]$. There are nine variables and one cointegration relationship between two of them. iv) $[9, 4]$. There are nine variables and three cointegration relationships between four of them.

For each of these four sub-DGPs, we perform the Johansen's trace test and the pairwise procedure. In the trace tests, we include only one lag, which is the true number. As discussed in §II.2.1.3, in the pairwise procedure, the lag structure depends on the *type of the pair* (i.e., both series have the common trend; one has it but the other does not; neither of the series has it). Thus, we try from one to five lags and select the optimal number according to the $AIC$ and the $BIC$.

Cointegration tests are made at 1% of significance, and the number of Monte Carlo replications was 1000. The experiments are performed with samples of 100, 200 and 400 data points.

The *Trace* columns in table II.1 contain the probabilities of finding the correct number of cointegration relationships by means of the Johansen's trace test, when all the $N$ variables are included in the model. The $PW$ columns contain the probabilities of finding cointegration in all the pairs that are truly cointegrated when the tests are done by the trace test, but in a pairwise fashion, and the lag length is selected according to the AIC (BIC). The preferred approach is marked in bold.

The table shows that nothing is lost by proceeding in a pairwise fashion. On the contrary, the

pairwise procedure outperforms the regular trace test. For large samples and a small number of series ($N$), both procedures provide the same results (which coincide with the theoretical ones). However, as the number of series increases or the sample size decreases, the differences in favor of the pairwise procedure become remarkable (bold entries are only in the PW columns). The largest differences are for the case with $N = 9$ and $T = 100$.

This result is closely related to those obtained by Lütkepohl et al. (2003) and Johansen (1995). These authors find that the power of cointegration tests decreases with the number of stochastic trends in the system, so that, for instance, it would be more difficult to detect a single cointegration relationship in a three-dimensional system than in a bivariate one. Note, however, that our result is not exactly the same since in table II.1 we are comparing estimation and testing cointegration in a single *full* model with several stochastic trends *vs.* doing it in several bivariate models (not one) with one stochastic trend. With our strategy we provide a better alternative to the Johansen's procedure for the case in which the cointegrated series constitute a fully cointegrated subset.

Note also the importance of lag selection for small sample sizes. The difference in the probabilities of finding all cointegration relationships with the pairwise procedure when $T = 100$ if we use the $AIC$ or the $BIC$ may be significant in favor of the latter. This is due to the efficiency losses generated by a larger number of regressors in small sample sizes (the $BIC$ tends to select shorter lag lengths).

Table II.1: Probability of finding all cointegration relationships. Comparison between the Trace test and the Pairwise procedure

| | | T = 400 | | | T = 200 | | | T = 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | PW | | | PW | | | PW | |
| N | $n_1$ | Trace | AIC | BIC | Trace | AIC | BIC | Trace | AIC | BIC |
| 6 | 2 | 0.97 | **0.99** | **0.99** | 0.91 | **0.99** | **0.99** | 0.65 | 0.92 | **0.96** |
| | 3 | **0.98** | **0.98** | 0.96 | 0.88 | **0.97** | 0.95 | 0.53 | 0.62 | **0.80** |
| 9 | 2 | 0.89 | **0.99** | **0.99** | 0.57 | **0.99** | **0.99** | 0.07 | 0.92 | **0.96** |
| | 4 | 0.95 | **0.98** | 0.94 | 0.67 | **0.92** | 0.91 | 0.38 | 0.39 | **0.64** |

Number of replications: 1000. *Trace* columns contain the probabilities of finding the correct number of cointegration relationships by means of the Johansen's trace test, when all the $N$ variables are included in the model. *PW* columns contain the probabilities of finding cointegration in all the pairs that are truly cointegrated when the tests are done by the trace test, but in a pairwise fashion and the lag length is selected according to the AIC (BIC) criteria.

### II.3.3 The behavior of the pairwise strategy

We now turn to the analysis of the pairwise strategy in the framework for which it has been designed: large $N$. As discussed in §II.2.1, the ideal procedure will: 1) from all the $N$ series, identify a large proportion of those that truly share the trend (those in $n_1$); and 2) not include wrong series in the estimated fully cointegrated subset(s), $\hat{n}_1$.

Conditions 1 and 2 are what Castle et al. (2011) call *potency* and *gauge*. While *gauge* measures the retention frequency of irrelevant variables when selecting among a (potentially large) set of candidates, *potency* denotes the average retention frequency of relevant variables.

For the two DGPs, we consider four scenarios. In all of them, we set $N = 100$, and they differ in the choice of $n_1$ — recall that we are using the notation $n_1$ both, to indicate the size of the 'fully cointegrated subset' and as its label. The four choices are $n_1 = 10$, $n_1 = 15$, $n_1 = 25$ and $n_1 = 40$. Scenarios 1 and 4 are motivated by results in Espasa and Mayo-Burgos (2013) about CPIs' components, and the other two scenarios are just to have intermediate structures. Additionally, we consider three possible sample sizes: $T = 100$, $T = 200$ and $T = 400$.

For each DGP, scenario and sample size, we perform 1000 Monte Carlo replications. In each replication we simulate a 100-dimensional model in which a subset of $n_1$ series share a single trend. The objective is to discover the series in $n_1$. To do that, we perform cointegration tests on all the 4950 bi-variate VAR sub-models that exist among the 100 series. Thus, for a particular DGP, scenario and sample size, we have 4.95 million sub-models (4950 for each replication). Since we have two DGPs, four scenarios and three sample sizes, we have $(2 \times 4 \times 3) \times 4.95 = 118.8$ million sub-models to estimate. Additionally, since the lag length for each of the 4950 sub-models of a particular replication is unknown, we select it with the AIC in a model with one cointegration relationship and admitting between one and five lags in the VEqM representation. The high complexity of the simulations led us not to consider more complex DGPs.

Since main conclusions do not change for the two basic DGPs, we focus on the results for *DGP 1*. Though in §II.3.4 we give some comments about the results for *DGP 2*, details are omitted and included in appendix A.3.

Let $Z_1$ be the number of correct series included in $\hat{n}_1$ (the estimated subset of variables that share a unique common trend). In addition to computing the potency, we compute $P(Z_1 \geq x)$ for $x \in [0, 1, ..., n_1]$. The larger this probability is for each $x$, the better the procedure's performance will be. Given the results presented in §II.2.1, we expect $P(Z_1 = n_1)$ to be close

to $1-\varphi$ in large samples.

Let $Z_2$ be the number of incorrect series included in the estimated fully cointegrated set. Apart from evaluating the gauge, we compute $P(Z_2 \geq x)$ for $x \in [1, ..., N-n_1]$. The larger this probability is, the worse the procedure's performance will be.

Figure II.1 includes the plots for $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for scenario 3 (since the basic conclusions do not change, plots for the other scenarios are not included in order to save space but are available upon request). Gauge and potency measures are delayed to §II.3.4, where a comparison of the pairwise strategy with an alternative DFM approach is carried out.

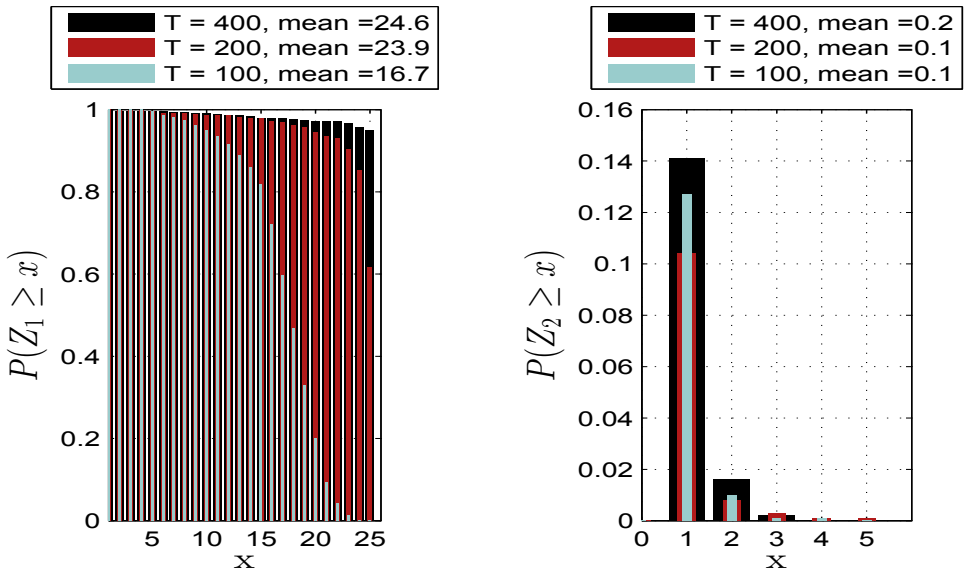



Figure II.1: Pairwise approach. $P(Z_i \geq x)$, Scenario 3 ($n_1 = 25$)

As figure II.1 shows, the procedure performs reasonably well for all sample sizes. For $T = 400$, the probability of including all of the 25 series is close to 99%. On average, for this sample size, we include 24.6 correct series in the fully cointegrated subset. This outcome is in line with theorem 1, which states that the asymptotic probability of finding cointegration in all the true cointegrated pairs is close to $(1-\varphi)$, with $\varphi$ being the nominal size of the tests.

On the other hand, the probability of including wrong series is quite low and fast decreasing with $x$. The probability of including at least one wrong series is 0.14. In other words, we include no wrong series in 86% of the experiments. Moreover, the probability of including more than two wrong series is close to zero.

On average, we include just 0.2 wrong series. Recall from §II.2.1 that an upper bound

for the expected number of wrong series in $\hat{n}_1$ is $E[W] = (N - n_1)\varphi$ (0.75 in scenario 3). Therefore, this result shows that the actual $E[W]$ is far from this upper bound, meaning that the assumption used to compute this bound is too permissive. This performance is appealing given that including wrong series is the worst possible error.

Importantly, we did not find any other fully cointegrated set composed by outsiders, in accordance with the DGP used to generate the data.

Finally, as figure II.1 shows, although gauge remains quite stable when the sample size varies, potency deteriorates as $T$ decreases. For instance, with $n_1 = 25$ (scenario 3, included in figure II.1), we go from $\bar{Z}_1 = 24.6$ with $T = 400$ to $\bar{Z}_1 = 16.7$ with $T = 100$. Still, in this case, we get a low gauge and capture 67% of the cointegrated series.

### II.3.4 Comparison with DFM

The third objective of the Monte Carlo experiments was to compare the pairwise procedure's performance with alternative approaches. In this section, we apply DFM and compare the results of both strategies.

For implementing the DFM, we apply the traditional Principal Components procedure to the whole data set and keep the number of factors suggested by the information criteria $IC_k$ and the three penalty functions detailed in Bai (2004). When each penalty function suggests a different number of factors, we choose the minimum; otherwise, we choose the mode. This procedure implies that we are not always using the same penalty function in each experiment, but it artificially helps the dynamic factors methodology to pick the correct number of factors (which is always one by construction). For the four scenarios, we proceed as suggested by Bai and Ng (2004): extracting the factors from the differenced data and integrating the results to obtain estimates of the original factors. This seems the most sensible procedure when $n_1$ is small compared to $N$. As Bai (2004) shows, computing the non-stationary dynamic factors from the levels of the variables is also a correct procedure under the assumption that the unique source of non-stationarity is the factors (idiosyncrasies must be stationary). This assumption ensures that the series are cointegrated and that the spurious regression problem will not be an issue. We also consider this option and, as expected, it is worse than extracting the factors from the differences, as all the series outside $n_1$ have non-stationary idiosyncrasies.

To analyze DFM's performance, we consider two criteria. First, an initial 'minimum quality

requirement' is that factors have to be cointegrated with the series in $n_1$. To analyze this issue, we perform Engle-Granger cointegration tests between each of the series in $n_1$ and each of the estimated factors. Note that the factors are not observed series, so a generated regressor problem may appear in these tests. The consideration of this problem would increase (in absolute value) the critical values of the tests, but since we are not dealing with it, we are being conservative in the sense that we will over-reject the null of no cointegration.

Second, in order to have a proper comparison with the pairwise procedure, we compute confidence bands for the factor loadings and identify those series with statistically significant factor loadings.[6] We consider the subset formed by these series the *DFM counterpart of the fully cointegrated subset*. Using those series, we can compute the gauge and potency of the DFM approach and the probability functions, $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$, that we used to study the performance of the pairwise approach.

Table II.2 contains the mean rejection frequencies of E-G tests at 10% of significance between the true series in $n_1$ and the estimated first factor for the four scenarios and different sample sizes. Figures in the table are averages across experiments and series in $n_1$.

For instance, for $T = 400$, in scenario 1, the average probability of rejecting the null of no cointegration between one of the ten series in $n_1$ and the estimated factor is 0.66. For $T = 100$, the traditional DFM procedure fails in the three scenarios, as the low figures in the first column indicate. Moreover, in scenario 1 ($n_1 = 10$), DFM also fails, even for moderately large samples ($T = 200$).

It is worth mentioning that these conclusions are not true if the Data Generating Process is a factor model instead of a VEqCM. In this case, all figures in table II.2 are close to one. This difference is due to the fact that the generated regressor problem is more important in *DGP 2* than in *DGP 1*. When the data come from a VEqM (DGP 1), the factor loadings of the corresponding DFM are $\beta_\perp$ (the orthogonal complement of $\beta$, see e.g., Escribano and Peña (1994)); thus, the factor loadings of the series in $n_1$ in the DFM that corresponds to *DGP 1* are equal to 1. In contrast, in *DGP 2*, factor loadings are taken from the uniform distribution with parameters $[0.1, 0.8]$. This makes the estimated factors noisier in *DGP 2* and the aforementioned generated regressor problem more relevant.

For space reasons, the probability density functions $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for the DFM

[6] We use the asymptotic variance of the factor loadings provided by Bai (2003).

Table II.2: Mean rejection frequencies in Engle-Granger tests between series in $n_1$ and the estimated factors from the whole data set. *DGP 1* ($H_0$ : *No cointegration*)

| Scenarios | T = 100 | T = 200 | T = 400 |
|---|---|---|---|
| 1 | 0.00 | 0.13 | 0.66 |
| 2 | 0.01 | 0.43 | 0.94 |
| 3 | 0.03 | 0.73 | 0.99 |
| 4 | 0.07 | 0.81 | 1.00 |

The null hypothesis is no cointegration.
Nominal size of the test: 10%.
Sce1: $n_1 = 10$, Sce2: $n_1 = 15$, Sce3: $n_1 = 25$, Sce4: $n_1 = 40$.

procedure are not included here. Instead, table II.3 includes the *gauge* and *potency* of the pairwise strategy and its DFM counterpart.

A general conclusion from this table is that the DFM procedure performs better than the pairwise approach in terms of *potency*. The probability of including a large proportion of the true series is greater when the DFM procedure is used (except for scenario 4 and $T = 200, 400$). However, *gauge* is substantially larger in scenarios 1 to 3, even for large sample sizes (except for scenario 3 and $T = 400$).

The DFM counterpart of the pairwise procedure fails to isolate the series in $n_1$ for scenarios 1 to 3. In scenario 1 with $T = 400$, on average, we include $0.081 \times (100 - 10) = 7.3$ wrong series. This bad performance substantially deteriorates as the sample size decreases. For T = 100, on average, 24.2 wrong series are included in $\hat{n}_1$. For scenarios 2 and 3, these comments are also valid in general, except for scenario 3 and $T = 400$.

Overall, the main conclusion from table II.3 may be that the pairwise procedure is preferred for situations of relatively small $n_1$ — a conclusion that seems more evident for relatively small sample sizes. When $n_1$ and $T$ become larger, DFM may be preferred. Note, however, that even in those situations — large $n_1$ and $T$ — the pairwise procedure also shows a very good performance.

These conclusions are also valid when the DGP is a Factor model instead of a VEqCM. Indeed, they seem even stronger as *gauges* of the DFM counterpart are larger (see table A.3 in appendix A.3). This difference is also related to the relationship between VAR models and DFM mentioned in relation to table II.2; the factor loadings of the DFM that are derived from *DGP 1* are substantially larger than those of *DGP 2*.

A final comparison between the two procedures could be their forecasting performance, but the one that performs better in grouping the components with common features is, in principle,

Table II.3: Comparison of gauge and gotency of the Pairwise procedure with its DFM counterpart

| | **Pairwise** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | | Sce 4 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 0.4 | 77.0 | 0.3 | 72.8 | 0.2 | 66.9 | 0.1 | 62.2 |
| T=200 | 0.3 | 96.9 | 0.2 | 96.5 | 0.2 | 95.6 | 0.1 | 94.9 |
| T=400 | 0.3 | 98.5 | 0.2 | 98.2 | 0.2 | 98.3 | 0.2 | 98.0 |
| | **DFM counterpart** | | | | | | | |
| | Sce 1 | | Sce 2 | | Sce 3 | | Sce 4 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 26.8 | 94.0 | 16.0 | 98.7 | 5.1 | 97.6 | 1.0 | 87.3 |
| T=200 | 18.3 | 99.7 | 6.7 | 99.9 | 0.7 | 99.4 | 0.0 | 93.0 |
| T=400 | 8.1 | 100.0 | 1.4 | 100.0 | 0.0 | 99.8 | 0.0 | 96.0 |

$Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
$Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
$Z_2$ = number of wrong series included in $\hat{n}_1$
$Z_1$ = number of correct series included in $\hat{n}_1$
$Nexp$ = number of experiments

expected to dominate the forecasting exercise.

The result that the DFM alternative performs badly for small $n_1$ was expected for the consistency of the estimated factors is based on the assumption of pervasiveness. Under this assumption, $1/N\Lambda'\Lambda$ (with $\Lambda$ being the matrix of factor loadings) tends to a positive definite matrix as $N$ goes to infinity (see e.g., assumption B in Bai (2003)). Since the meaning of pervasiveness in finite samples is not clear, the usefulness of the experiments above is to show that, for the data structures we are interested in, factors can be considered to be non-pervasive, what leads the principal components strategy to fail.

Instead of the principal components strategy, we could have used some maximum likelihood approach applicable in large dimensional frameworks ( see e.g., Doz et al. (2012) and Doz et al. (2011)). As these approaches also rely on the assumption of pervasiveness, they will also show difficulties when $n_1$ is small.

To close this section, it is worth highlighting an apparent contradiction between the relatively good performance of DFM in scenario 4 for $T = 100$, showed in table II.3, and its bad performance in terms of cointegration, shown in table II.2. According to table II.3, the procedure succeeds in identifying the correct series in $n_1$, giving significant loadings to the correct series

and insignificant ones to the series outside $n_1$. However, according to table II.2, the estimated factor (that aggregates the series using the estimated loadings) is not cointegrated with the series in $n_1$, which are precisely those whose loadings are significant.

This is not necessarily a contradiction. Although for $T = 100$, the procedure, on average, correctly identifies a large proportion (87.3%) of the 40 series, the remaining loadings are substantially more noisy than for $T = 400$. This explains the apparent contradiction between Tables table II.2 and II.3, and suggests — in line with Boivin and Ng (2006) and Beck et al. (2015) — that results could be improved by extracting the factors only from the series that show significant loadings (at least for $T = 100$).

## II.4 Preliminary discussion for the of $N$ going to infinity

Up to now, the asymptotic analysis was for the case of 'large' but fixed $N$, and $T$ going to infinity. The pairwise procedure has the advantage of being able to deal both with large and small $N$ without the need of requiring $N$ going to infinity. This is, in our opinion, a strength, not a weakness, as assuming that the number of components of an aggregate goes to infinity may be unrealistic in many situations. However, it is important to study how the procedure behaves as $N$ goes to infinity. In this section we give some initial insights about this issue.

We first study the problem of including wrong series in the estimated fully cointegrated subsets when $N$ goes to infinity. Next, we focus on the 'discovery' of non-existent subsets. The inclusion of all the correct series is covered by theorem 1.

### II.4.1 Inclusion of wrong series in the estimated subsets

As argued in §II.2.1.2, to include a wrong series in an estimated fully cointegrated subset we need to find cointegration with all the series of the subset, and the probability of such an event can be written as:

$$P(WR_1 \cap ... WR_{n_1}) = P(WR_1|WR_2, ..., WR_{n_1}) \times P(WR_2|WR_3, ..., WR_{n_1}) \times ... P(WR_{n_1-1}|WR_{n_1}) \times P(WR_{n_1}), \tag{II.20}$$

where, under assumption 2, $WR_i$ is the event of wrongly rejecting $r = 0$ between a series outside a fully cointegrated subset and the $i^{th}$ series of that subset. Since for $T \to \infty$ $P(WR_{n_1}) \to \varphi$

(the nominal size of the individual tests), the expected proportion of wrong series will be:

$$\frac{(N-n_1)P(WR_1|WR_2,...,WR_{n_1})\times ...\times P(WR_{n_1-1}|WR_{n_1})\times\varphi}{n_1}. \tag{II.21}$$

As we also argued in §II.2.1.2, in the extreme case that all conditional probabilities in the numerator of eq. (II.21) are equal to one, the expected proportion of wrong series would be $\frac{(N-n_1)\varphi}{n_1}$. In our simulation experiments (see §II.3.3) we found that the proportion of wrong series is far below that quantity, meaning that there should be some conditional probabilities smaller than one.

Assume that there is a proportion $\delta$ of the $n_1-1$ conditional probabilities that do not exceed a fixed threshold $\varphi_{max}$, with $\varphi_{max}<1$. Then, an upper bound for the expected proportion of wrong series will be:

$$\overline{E[W]}=\frac{(N-n_1)\varphi_{max}^{\delta(n_1-1)}\varphi}{n_1} \tag{II.22}$$

If $n_1$ is fixed, when $N$ goes to infinity, eq. (II.22) also goes to infinity. For avoiding this, we also need $n_1\to\infty$. But note that we do not need $N$ and $n_1$ to grow at the same rate. Indeed, if in addition to $\varphi_{max}$, $\delta$ is also fixed, a sufficient condition for avoiding eq. (II.22) going to infinity is $\frac{n_1}{log(N)}\to c$, for some positive $c$.

**Proof** *First, take logs in eq. (II.22):*

$$log(\overline{E[W]})=log(N-n_1)+(n_1-1)\delta log(\varphi_{max})+log(\varphi)-log(n_1),$$

*Now let* $n_1=cN^{1/\gamma}$*, with* $\gamma>1$ *and* $c>0$*, then:*

$$log(\overline{E[W]})=log(N-cN^{1/\gamma})+(cN^{1/\gamma}-1)\delta log(\varphi_{max})+log(\varphi)-log(cN^{1/\gamma}) \tag{II.23}$$

*Let* $log(\overline{E^*[W]})=log(N)-c^*N^{1/\gamma}$*, with:*

$$c^*=-c\delta log(\varphi_{max}). \tag{II.24}$$

*Since* $\varphi_{max}$ *is fixed, for sufficiently large* $N$*,* $log(\overline{E[W]})\leq log(\overline{E^*[W]})$*. Then, having a constant* $\overline{E^*[W]}$ *is a sufficient condition for keeping* $\overline{E[W]}$ *lower than a certain threshold.*

*To find the condition for constant* $log(\overline{E^*[W]})$ *write;*

$$log(N) - c^* N^{1/\gamma} = C, \tag{II.25}$$

*from where;*

$$\gamma = \frac{log(N)}{log[log(N) - C] - log(c^*)}, (if\ c^*\ does\ not\ go\ to\ 0) \rightarrow \frac{log(N)}{log[log(N)]} > 1. \tag{II.26}$$

*Noting that* $N^{log(log(N))/log(N)} = log(N)$*, completes the proof.* ■

Since we assumed that $\varphi_{max}$ is fixed, the condition '$c^*$ *does not go to 0*' is equivalent to '$\delta$ *does not go to zero*'. Note however that even when $\delta \rightarrow 0$, we can still have $\gamma < 1$ and not increasing $\overline{E[W]}$. The condition for $\gamma > 1$ and not increasing $\overline{E[W]}$ would be:

$$\frac{log(N)}{log[log(N)] - log(c^*)} > 1 \Leftrightarrow \frac{N}{c^*} > 1 \Leftrightarrow \frac{N}{\delta} > 1, \text{ as } N \rightarrow \infty \tag{II.27}$$

Meaning that $1/\delta$ can grow, at most, at the same rate as $N$.

Although we do not have a formal proof to show that II.27 holds, we perform a simple Monte Carlo study that suggests it does.

#### II.4.1.1 Monte Carlo study

To evaluate the arguments made above we perform two simulations experiments. The DGP is the VAR of §II.3.1 and the number of replications is 500. In each experiment we apply the pairwise procedure for discovering fully cointegrated subsets and count the number of variables for which we wrongly rejected the null of $r = 0$ at a significance level of 0.1 with *all* the variables that truly belong to $n_1$. While in the first experiment we maintain constant the ratio $n_1/log(N)$ as $N$ increases, in the second one the constant ratio is $n_1/N$.

**Experiment 1.** We simulate five different scenarios for $N$ and $n_1$: i) $N = 150$, $n_1 = 100$; ii) $N = 303$, $n_1 = 114$; iii) $N = 610$, $n_1 = 128$; iv) $N = 911$, $n_1 = 136$; and, v) $N = 1503$, $n_1 = 146$.

**Experiment 2.** We simulate four different scenarios for $N$ and $n_1$: i) $N = 50$, $n_1 = 5$;

ii) $N = 100$, $n_1 = 10$; iii) $N = 200$, $n_1 = 20$; iv) $N = 400$, $n_1 = 40$; v) $N = 800$, $n_1 = 80$; and vi) $N = 1200$, $n_1 = 120$;

Call the cases in which the null of $r = 0$ between a series outside $n_1$ and *all* the elements inside $n_1$ was wrongly rejected as *wrong inclusions*. The results for the ratios *wrong inclusions*$/n_1$ are summarized figure II.2 and figure II.3, for experiment 1 and 3, respectively. While figure II.2 suggests that $\delta$ is not constant, figure II.3 suggests that condition II.27 holds. This mean that although $n_1$ has to grow at a faster rate than $log(N)$, it can grow slower than $N$, so we may have $N/n_1 \to \infty$.

The pervasiveness assumption in large dimensional factor models requires that $1/N\Lambda'\lambda$ tends to a positive definite matrix, where $\Lambda$ contains the factor loadings (see e.g., assumption B in (Bai, 2003)). For the case of only one factor, this assumption implies that the number of non-zero factor loadings has to grow at the same rate as $N$. Therefore, our result that $N/n_1$ can go to infinity means that we do not need pervasiveness.

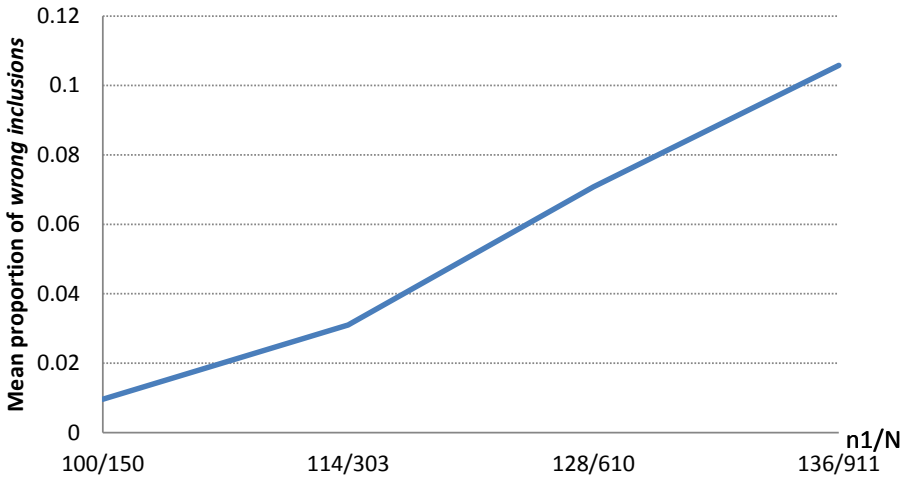



Figure II.2: Mean proportion of *wrong inclusions* in the estimated fully CI subset as $N$ increases and $log(N)/n_1$ remains constant

### II.4.2 False discovery of fully cointegrated subsets

In appendix A.2 we argued that the independence assumption is sensible for series outside fully cointegrated subsets. Then, we can use the result of Matula (1976) about the size of maximum fully connected sub-graphs. For our case of interest, and under assumption 2, the size of this fully connected sub-graph has a strong peak at $2log(N-n_1)/log(1/\varphi)$, with $\varphi$ being the nominal size of the individual tests. Therefore, in order to avoid this value to go to infinity, we should

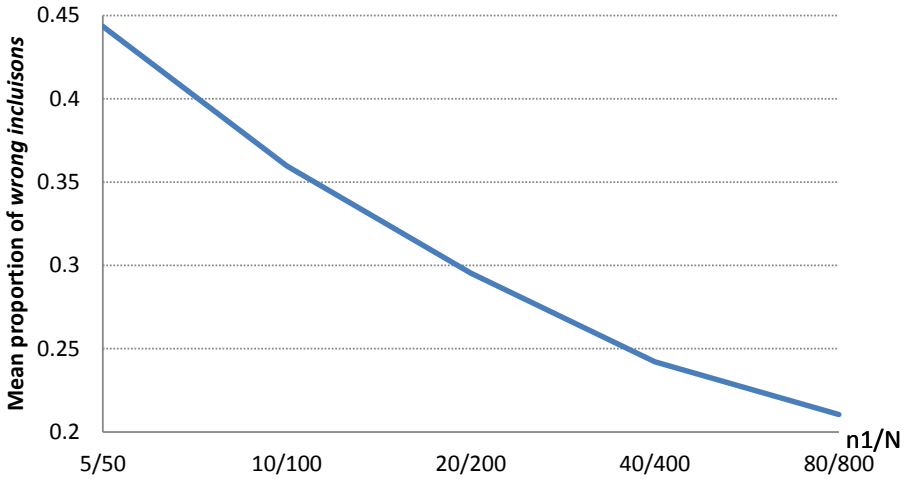



Figure II.3: Mean proportion of *wrong inclusions* in the estimated fully CI subset as $N$ increases and $N/n_1$ remains constant

use a significance level of, say, $N^{-1/\lambda}$, for some $\lambda > 0$. In this case, the size of the maximum false fully cointegrated subset will be $2\lambda$. For being able to use such a significance level, we need $T$ to grow at a rate larger than or equal to $N^{1/\lambda}$. Since $\lambda$ can be larger than 1, we can deal with the case of $N/T \to \infty$.

## II.5 Concluding Remarks

In this paper, we studied the properties of a pairwise procedure for testing cointegration between all the possible pairs of an aggregate's components at the maximum level of disaggregation. This procedure allows us to discover blocks of series that share a unique common trend (*fully cointegrated block*). The main theoretical result is that cointegration tests inside those blocks are asymptotically equivalent, in the sense that the probability that all tests deliver the same conclusion is close to 1 as $T$ goes to infinity independently of the number of series. Thus, multiple testing is not an issue for pairs of components inside a fully cointegrated block.

In a Monte Carlo experiment, we confirmed the asymptotic results and compared the performance of the pairwise approach with that of a DFM alternative. This comparison showed that the pairwise procedure dominates in situations in which the number of series that share the trend ($n_1$) is relatively small with respect to the total number of components, $N$. The DFM alternative fails in those situations. For relatively large fully cointegrated blocks and $T$, the DFM alternative may be preferred, though the pairwise approach also performs very well in those scenarios.

Other relevant results are those related to the specification of the bivariate sub-models in which cointegration tests are performed. First, we argued that changing the normalization of the cointegrating matrix may lead to an improvement in the forecasting accuracy of the individual models for the components. This potential improvement is derived from a possible reduction in the number of regressors — which does not cause information losses — and, hence, in the estimation uncertainty of those models. Second, we established the conditions under which the forecasting strategy in Espasa and Mayo-Burgos (2013) (who, in the single-equation models, include just one cointegration restriction with respect to the sub-aggregate of the fully cointegrated components) is correct. As the conditions are quite stringent, we proposed a generalization of that strategy that does not rely on them. Our proposal is to consider all the $n_1 - 1$ cointegration relationships as potential regressors for the forecasting equations of fully cointegrated components and to select the significant ones by applying a selection algorithm.

Finally, we found that, in moderately short samples, the pairwise strategy leads to power improvements with respect to a regular Johansen's test applied to a (reduced) group of series that share a common trend. As shown in table II.1, these improvements are remarkable in many situations.

### II.5.1 Discussion

To close this chapter, a comment about our testing strategy for discovering common trends is worthwhile. The usual approach of the Johansen's test is to perform sequential tests, starting with $r = 0$ vs. $r > 0$, and continuing up to the first no-rejection. Asymptotic power equal to one ensures that in infinitely large samples the probability of finding $r < r^*$ is zero (with $r^*$ being the true cointegration rank).

For cointegrated pairs ($r^* = 1$) this ensures that finding $r = 0$ is not an issue. Therefore, for these pairs, the problem could be wrongly reject the null of $r = 1$ in favor of $r = 2$ (type 1 error). The probability of this event is the nominal size of the test ($\varphi$). Given the large amount of test we are doing, the probability of not making the type 1 error in any of the truly cointegrated pairs would be close to zero if the tests were independent. This means that the probability of discovering the 'true' fully cointegrated subset would be close to zero. Fortunately, far from being independent, cointegration tests inside a *fully cointegrated subset* are asymptotically equivalent (theorem 1), and the probability of finding $r = 1$ in all the pairs is close to $1 - \varphi$. Thus, there is not a multiple testing issue for this pairs.

For pairs that are not cointegrated ($r^* = 0$) there is a probability $\varphi$ of finding $r = 1$, what could lead to include wrong elements in the estimated *fully cointegrated subset*. However, the 'full cointegration' requirement protects us against this problem, as the null has to be wrongly rejected with all the series in the estimated subset. For $\varphi = 0.01$ and $N = 100$ we could expect to include, at most, one wrong series.

Instead of performing cointegration tests we could have tested for common trends, and this could have changed our procedure's properties.

The election of the Johansen's procedure is based on Gonzalo (1994), who compares five different alternatives for estimating long-run equilibrium relationships (including the principal components approach in Stock and Watson (1988)) and find that the Johansen's procedure has the best properties. Results in Gonzalo (1994) are for the estimation of the cointegration relationships, not for determining the number of common trends. Thus, we could think of estimating the cointegration relationships by maximum likelihood and then test for common trends using the Stock and Watson (1988) test.

When testing for common trends the contrast is $k$ vs. $m$ common trends with $k > m$. Therefore, asymptotic power equal to one ensures that we will never find $ct > ct^*$ (where $ct^*$ is the true number of common trends).

For testing for common trends in the pairwise approach we could proceed as follows: first test $ct = 2$ vs. $ct = 1$, next, if the null is rejected, test $ct = 1$ vs. $ct = 0$. This is the same sequence as that of the Johansen procedure, for $ct = 2$ is equivalent to $r = 0$, $ct = 1$ is equivalent to $r = 1$, and $ct = 0$ is equivalent to $r = 2$. Therefore, the analysis keeps exactly the same. For cointegrated pairs ($ct^* = 1$), asymptotic power ensures that the first null will always be rejected, thus finding $ct = 2$ ( $r = 0$ ) is not an issue. When testing $ct = 1$ vs. $ct = 0$ ($r = 1 vs. r = 2$) there is a probability $\varphi$ of wrongly rejecting the null. If tests are independent the probability of not making any wrong rejection will be close to zero. However, theorem 1 will probably work for common trends tests also.

For not cointegrated pairs ($ct^* = 2$), there is a probability $\varphi$ of wrongly rejecting the null and finding $ct = 1$ ($r = 1$). This could lead to include wrong elements, but the full cointegration requirement will protect us against this problem.

# Appendix A

# Appendix to Chapter II

## A.1 Testing the null of a unique common trend in a large group of series

Given a group of $N$ series, theorem 1 allows allows us to look for subsets of series that share a single common stochastic trend in a pairwise fashion — that is, to perform cointegration tests between all possible pairs of series and group all the series that are pairwise cointegrated. Pesaran (2007) poses another interesting question: i.e., whether or not the $N$ series at hand share a single stochastic trend. In this subsection, we argue how one might proceed to tackle this question.

In what follows, we simulate critical values for the null that the number of cointegration relationships in *all* the pairs among a subset of $n_1$ series is 1 vs. the alternative that at least in one of them it is 2. We denote this test $-2lnQ(JH(1)/JH(2))$ — the $J$ standing for *'Joint'*.

As Johansen (1995) shows, the test statistic for the null of zero versus $k$ cointegration relationships — denoted as $-2ln(H(0)/H(k))$ — in a $k$ dimensional model converges weakly to:

$$tr\{\int_0^1 (dB)B'[\int_0^1 BB'du]^{-1}\int_0^1 B(dB')\},$$

where $B$ is a standard Brownian motion. Johansen (1995) also shows that the asymptotic distribution of $-2lnQ(H(r)/H(k))$ is the same as $-2lnQ(H(0)/H(k-r))$. Given this result, the author simulates asymptotic tables for $k-r$ dimensional random walks. Note, however, that this strategy is not useful for the purposes of this section because we need to keep the same

single common trend for all the pairs. Thus, we use a different approach.

Consider the following $DGP$: $\Delta X_t = c_0 + d_0 t + \alpha(\beta' X_{t-1} + c_1 + d_1 t) + \epsilon_t$, where $\epsilon_t \sim N(0, I_2)$, $\alpha = [0, 1]'$, $\beta = [1, -1]'$ and $X_0$ is fixed at $[0, 0]'$. We simulate this $DGP$ for three different assumptions of the deterministic terms: **Case 1:** $c_0 = d_0 = c_1 = d_1 = 0$. **Case 2:** $c_0 = 10$, $d_0 = c_1 = d_1 = 0$. **Case 3:** $c_0 = 10$, $c_1 = [1, 0.7]'$, $d_0 = d_1 = 0$. Although different from Johansen's proposal, with this strategy, original critical values can be reproduced (results are available upon request).

The number of experiments is 5000. For each experiment, we simulate 5000 pairs with the same original stochastic trend — i.e., $\sum_{t=1}^{T} \epsilon_{1,t}$, with $T = 1000$. Thus, for each of the 5000 experiments, we have 5000 test statistics.

A necessary and sufficient condition for not rejecting the null of $r = 1$ $(k - r = 1)$ $vs$ $r = 2$ $(k - r = 0)$ for all the pairs, is to not reject it for the pair with the maximum test statistic. Therefore, the test statistic for the null that all the pairs are cointegrated is the maximum of all the pairs considered.

In order to assess the magnitude of the difference between the quantiles of $-2lnQ(JH(1)/JH(2))$ and those of the original Johansen's distribution, Table A.1 includes the cumulative probabilities of $-2lnQ(JH(1)/JH(2))$ at Johansen's quantiles. The slight differences between these probabilities and Johansen's confirm that both distributions are very close, even when considering 4950 pairs.

Therefore, for testing the null hypothesis that a given group of series share a unique common trend, we can first test $r = 0$ vs $r > 0$ in all possible pairs using regular critical values — asymptotic power is 1. Then, if all hypothesis are rejected, test $r = 1$ vs $r = 2$ in all pairs, store the maximum test statistic, compare it to the regular critical values, and reject the null of a unique common trend if the maximum statistic is larger than the critical value. Asymptotically, this procedure delivers the correct size.

Table A.1: Quantiles comparison. Cumulative probabilities of $-2lnQ(JH(1)/JH(2))$ at Johansen's quantiles

| | **50%** | **75%** | **80%** | **85%** | **90%** | **95%** | **97.5%** | **99%** |
|---|---|---|---|---|---|---|---|---|
| **Model: No deterministic terms ($c_0 = c_1 = d_0 = d_1 = 0$)** | | | | | | | | |
| $n_1$=5 (10 pairs) | 0.48 | 0.74 | 0.79 | 0.84 | 0.89 | 0.95 | 0.97 | 0.99 |
| $n_1$=10 (45 pairs) | 0.47 | 0.73 | 0.78 | 0.83 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=50 (1225 pairs) | 0.46 | 0.71 | 0.77 | 0.82 | 0.88 | 0.94 | 0.96 | 0.98 |
| $n_1$=100 (4950 pairs) | 0.45 | 0.71 | 0.76 | 0.82 | 0.87 | 0.94 | 0.96 | 0.98 |
| **Model: Restricted constant ($c_1 = d_0 = d_1 = 0$)** | | | | | | | | |
| $n_1$=5 (10 pairs) | 0.48 | 0.74 | 0.79 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=10 (45 pairs) | 0.47 | 0.73 | 0.79 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=50 (1225 pairs) | 0.46 | 0.71 | 0.78 | 0.83 | 0.88 | 0.94 | 0.97 | 0.99 |
| $n_1$=100 (4950 pairs) | 0.46 | 0.71 | 0.78 | 0.83 | 0.88 | 0.94 | 0.96 | 0.99 |
| **Model: Unrestricted constant ($d_0 = d_1 = 0$)** | | | | | | | | |
| $n_1$=5 (10 pairs) | 0.49 | 0.74 | 0.79 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=10 (45 pairs) | 0.49 | 0.73 | 0.79 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=50 (1225 pairs) | 0.48 | 0.73 | 0.78 | 0.83 | 0.89 | 0.94 | 0.97 | 0.99 |
| $n_1$=100 (4950 pairs) | 0.48 | 0.73 | 0.78 | 0.83 | 0.88 | 0.94 | 0.97 | 0.99 |

## A.2 False rejection of $r = 0$ when none of the series belongs to the fully cointegrated subset

Using the arguments in the proof of theorem 1 it can be easily seen that the elements of $\underset{\sim}{\Psi}$ are $Op(T)$, and that the matrices $S_{11}$ and $S_{11}^*$ differ in $Op(T)$ terms, even when comparing the tests statistics for two pairs with one series in common. This implies that the dependence among the tests outside fully cointegrated subsets will be low, even when considering the pairs formed by a single series and all the others. A simple Monte Carlo study confirms this statement.

If the tests statistics are independent, their results can be considered as random graph with edge probability $\varphi$. Therefore the expected number of fully connected subsets of size $k$ is $C_k^{N-n_1}\varphi^{k(k-1)/2}$, which goes to zero very fast, where $n_1$ is the size of the fully cointegrated subset, and $N$ the total number of series.

Matula (1976) showed that the size of the maximal clique[1] in a random graph with $N$ edges and edge probability $\varphi$ converges in probability $2ln(N)/ln(1/\varphi)$ as $N$ goes to infinity. Therefore, disregarding estimated fully cointegrated subsets of size less than or equal to $2ln(N)/ln(1/\varphi)$

[1]The maximal sub-graph in which all nodes are pairwise connected.

constitutes a strong protection against finding subsets between series outside $n_1$.

### A.2.1 Monte Carlo study

We generate $N+1$ random walks, pick one of them and perform Johansen's cointegration tests between the selected series and the remaining $N$. Call these tests $-2lnQ(H^*(0)/H^*(1))$. We replicate this experiment 1000 times. The DGP is:

$$Y_{t_{(N+1)\times 1}} = Y_{t-1} + e_t, \tag{A.1}$$

with $e_t \sim N(0, \Sigma)$. We set $N = 1000$, and the sample size is $T = 400$. The structure of $\Sigma$ is not relevant. We consider $\Sigma = I$, and $\Sigma = 0.95(\mathbb{1} - I) + I$, where $\mathbb{1}$ is a $(N+1) \times (N+1)$ matrix full of ones. This second option is a matrix with ones in the main diagonal and 0.95 elsewhere.

If the $N$ tests statistics of each replica where independent, they should be follow the Johansen's distribution. To assess if this is the case, for each replica, we compute the cumulative probability at the Johansen's quantiles and take the mean and the median across experiments.

Table A.2: Quantiles comparison. Cumulative probabilities of $-2lnQ(H^*(0)/H^*(1))$ at the Johansen's quantiles.

| | **50%** | **75%** | **80%** | **85%** | **90%** | **95%** | **97.5%** | **99%** |
|---|---|---|---|---|---|---|---|---|
| | | | | **$\Sigma = I$** | | | | |
| Mean | 0.48 | 0.73 | 0.79 | 0.84 | 0.89 | 0.94 | 0.97 | 0.99 |
| Median | 0.53 | 0.79 | 0.84 | 0.89 | 0.93 | 0.97 | 0.98 | 0.99 |
| | | | | **$\Sigma = 0.95(\mathbb{1} - I) + I$** | | | | |
| Mean | 0.47 | 0.72 | 0.78 | 0.83 | 0.88 | 0.94 | 0.97 | 0.99 |
| Median | 0.50 | 0.80 | 0.85 | 0.90 | 0.94 | 0.98 | 0.99 | 0.99 |

The slight differences between the cumulative probabilities of $-2lnQ(H^*(0)/H^*(1))$ the those of the Johnasen's distribution confirm that the assumption of independence outside $n_1$ is sensible.

## A.3 Simulation results when the DGP is a DFM

Table A.3: Comparison of gauge and potency of the Pairwise procedure with its DFM counterpart. DGP is a DFM

| | **Pairwise** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | | Sce 4 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 0.2 | 91.8 | 0.1 | 90.3 | 0.1 | 89.5 | 0.1 | 88.5 |
| T=200 | 0.2 | 96.2 | 0.2 | 95.5 | 0.2 | 95.2 | 0.2 | 94.6 |
| T=400 | 0.4 | 94.8 | 0.3 | 94.2 | 0.3 | 93.7 | 0.2 | 93.1 |
| | **DFM counterpart** | | | | | | | |
| | Sce 1 | | Sce 2 | | Sce 3 | | Sce 4 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 47.5 | 92.5 | 32.3 | 92.0 | 13.1 | 88.0 | 2.5 | 78.2 |
| T=200 | 38.5 | 96.6 | 20.3 | 95.1 | 4.5 | 90.3 | 0.2 | 79.3 |
| T=400 | 26.4 | 98.3 | 9.8 | 96.8 | 0.8 | 91.8 | 0.0 | 79.6 |

$Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
$Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
$Z_2$ = number of wrong series included in $\hat{n}_1$
$Z_1$ = number of correct series included in $\hat{n}_1$
$Nexp$ = number of experiments

*This page intentionally left blank.*

# Chapter III

# Modeling a large set of disaggregates with common trends and outliers

## III.1 Introduction

The general practice of empirical macroeconomic analysis is to focus just on aggregated variables such as the CPI, the GDP, the Industrial Production Index (IPI), Imports, Exports, Unemployment, etc. As all these variables are constructed by aggregating their components, limiting the analysis just to the aggregate implies denying a great amount of information. The starting point of this chapter, is the same as that of the previous one; this information should not be disregarded at the outset because it may be of interest for decision makers.

In this chapter we expand and apply the pairwise approach studied in chapter II. Our application is devoted to model and forecast all the components of the US CPI. The procedure in chapter II consists of performing Johahnsen's cointegration tests between all the $N(N-1)/2$ pairs that exist among the $N$ components of the CPI, and then, building subsets of components such that, in each subset, all its elements are pairwise cointegrated. Those subsets are denoted as *fully cointegrated* and have the property that all the components inside them share a unique common trend. In chapter II we showed the good performance of the procedure to *discover* the true *fully cointegrated groups* when they exist. For forecasting the components, in line with Espasa and Mayo-Burgos (2013), in chapter II we proposed to use single equation models for all of them including as potential regressors the cointegration relationships (when it is the case), own lags, and lags of the other components or some intermediate sub-aggregates. The selection among all these potential regressors can be done with the automatic selection al-

gorithm *Autometrics* (see Doornik (2009)). This strategy was commented but not implemented in chapter II.

As a global measure of the accuracy of the individual forecasts Espasa and Mayo-Burgos (2013) propose to asses the accuracy of the indirect forecast of the CPI obtained by adding up the disaggregated forecasts. Considering also common cycles restrictions, but using simpler approaches for cointegration tests and for specifying the single equation models than those in chapter II, Espasa and Mayo-Burgos (2013) show that the indirect forecasts for the CPI in US, UK and Euro Area outperforms both, direct procedures and other indirect competitors.

Our extensions to the proposal of chapter II are intended to robustify the procedure when applied to real data which may be subject to irregularities and/or to short samples problems. The robustification concerns both, the construction of the *fully cointegrated subsets* and the estimation of the single equation models.

Regarding the *fully cointegrated subsets*, in chapter II we showed that the potency of our procedure for discovering the subsets suffers an important deterioration in small samples. In this chapter we propose a small samples correction and study its properties by Monte Carlo. Additionally, as it is shown by several authors (see e.g, Johansen et al. (2000), Saikkonen and Lutkepohl (2002) and Nielsen (2004)), cointegration tests are strongly affected by the presence of outlying observations. For dealing with this issue we propose a solution that combines the feasible GLS approach in Saikkonen and Lütkepohl (2000) and Saikkonen and Lutkepohl (2002) with the Impulse Indicator Saturation (IIS) methodology in Santos et al. (2008), and study its performance by Monte Carlo.

Apart from robusfying the search of cointegration relationships against data irregularities and short samples issues, we extend the search by the inclusion of weakly exogenous variables and the admission of more general short run dynamics as suggested by Aron and Muellbauer (2013). Additionally, we add some 'quality' controls to the estimated cointegration relationships to be accepted as 'good' ones.

The rest of the chapter is organized as follows. In §III.2, we describe the outlier correction procedure and study its properties by Monte Carlo. §III.3 is devoted to the small samples correction procedure. In §III.4, we include the modifications of the cointegration relationships search, and in §III.5 we give a detailed description of the pairwise algorithm enlarged with our contributions. Finally, §III.6 and §III.7 include the application to the US CPI and the

conclusions, respectively.

## III.2 Data irregularities

The presence of outlying observations can generate devastating effects on parameter estimates and inferential conclusions if not adequately treated. Dealing with this issue in non-stationary data is specially troublesome since results on unit root and cointegration rank tests are affected by the presence of outliers and breaks (location shifts), and the other way round: tests for the presence of outliers and breaks will also be affected by the presence of unit roots (see *inter alia* Perron (1989); Perron and Vogelsang (1992); Doornik et al. (1998); Johansen et al. (2000), Perron and Rodríguez (2003), Perron (2006) and Juselius (2006)).

When trying to distinguish between a unit root and a (trend-) stationary process, traditional tests will tend to keep the null of unit root when the process suffers location shifts but is stationary within regimes. Additionally, when trying to detect a location shift, most tests will reject the null of no break when the process has a unit root with constant parameters. Similarly, as noted by Quintos (1998) and stressed by Perron (2006), tests for location shifts on cointegrated systems will over-reject the null of no break when the cointegrating rank is over specified (when the number of unit roots in the system is under-specified). Furthermore, cointegration rank tests will under estimate the number of cointegration relationships if the data is subject to location shifts. Therefore, a circular problem exists when dealing with non-stationary series that may be subject to location shifts.

On the other hand, an additive outlier (AO) has the opposite effect on unit root tests. As noted by Franses and Haldrup (1994), the presence of AOs induce a negative MA component in the residuals making traditional unit root tests to over-reject the null of unit roots.

These facts make the assessment of cointegration rank (and/or integration order) in the presence of outliers and breaks to be difficult because the appropriate treatment of these observations and the cointegration rank should, in principle, be decided simultaneously.

Although the pernicious effects of outlying observations in cointegration analysis is very well documented in the literature, the question of how to deal with these issues has not clear and generally accepted answer. In this section we propose an empirical strategy for dealing with these issues in the framework of the pairwise approach.

### III.2.1 Two strategies for dealing with outlying observations

#### III.2.1.1 Feasible GLS procedure

As analyzed by Johansen et al. (2000), in the traditional Gaussian approach the asymptotic distribution of the cointegration rank test changes due to the presence of structural breaks in the constant or the deterministic trend, and the new distribution depends on the breaks' dates. These new distributions can be approximated by *Gamma* functions whose parameters (mean and variance) can also be approximated by certain functions of the number of non-stationary relations and the location of break points. Johansen et al. (2000) restrict their attention to the case of a broken level in a model without deterministic trend, and a broken linear trend in a model with linear trend only outside the cointegration relationship, in both cases they consider a maximum of two breaks.

The feasible GLS estimation procedure of the coefficients associated to the deterministic parameters proposed by Saikkonen and Lütkepohl (2000) and Saikkonen and Lutkepohl (2002) has the virtue that test statistic's distribution does not depend on the break dates. The authors propose a two step procedure for dealing with deterministic components and interventions in the cointegrated VAR model. Their starting point is the following unobserved components model for the $N$-dimensional vector $X_t$:

$$X_t = \mu_0 + \mu_1 t + \theta DS_t + Y_t, \quad t = 1, 2, ..., \tag{III.1}$$

where $DS_t = 0$ if $t < T_1$, $DS_t = 1$ if $t \geq T_1$, and it is assumed that $\lambda = T_1/T$ remains fixed as $T$ grows. $Y_t$ is and $N-$dimensional unobserved vector, assumed to be at most I(1) and to follow a cointegrated VAR(p) process, whose VEqCM representation is: $\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta Y_{t-j} + \epsilon_t, \quad t = 1, 2, ...,$ where $\alpha$ and $\beta$ are $N \times r$ matrices; with $0 < r < N$; $r$ being the number of cointegration relationships; and $\epsilon_t$ is a Gaussian white noise. Expressing eq. (III.1) in terms of observable variables we get:

$$\begin{aligned} \Delta X_t &= \upsilon + \alpha(\beta' X_{t-1} - \vartheta(t-1) - \kappa DS_{t-1}) + \textstyle\sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \sum_{j=1}^{k-1} \Upsilon_j \Delta DS_{t-j} + \epsilon_t \\ &= \upsilon + \Pi^* X^*_{t-1} + \textstyle\sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \sum_{j=1}^{k-1} \Upsilon_j \Delta DS_{t-j} + \epsilon_t, \quad t = k+1, k+2, ..., \end{aligned} \tag{III.2}$$

where $\upsilon = -\alpha\beta'\mu_0 + \Psi\mu_1$ (with $\Psi = I_n - \Phi_1 - ... - \Phi_{k-1}$), $\vartheta = \beta'\theta$; $\Phi^* = \alpha[\beta' : \vartheta : \kappa]$;

$X_{t-1}^{*\prime} = [X_{t-1}', -(t-1), -DS_{t-1}]$; and $\Omega$ is the covariance matrix of $\epsilon$. Finally:

$$\Upsilon_j = \begin{cases} \theta, \ j = 0 \\ -\Phi_j \theta, \ \ j = 1, ..., k-1 \end{cases} \tag{III.3}$$

Saikkonen and Lütkepohl (2000) propose to obtain initial estimators $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\Phi}_j$ and $\tilde{\Omega}$ from eq. (III.2) without considering the restrictions in eq. (III.3) and using the rank $r_0$ that is specified in the null hypothesis of the cointegration rank test. Then, defining $\tilde{A}(L) = I_n\Delta - \tilde{\alpha}\tilde{\beta}'L - \tilde{\Phi}_1\Delta L - ... - \tilde{\Phi}_{k-1}\Delta L^{k-1}$, and $\tilde{Q}$ such that $\tilde{Q}\tilde{Q}' = \tilde{\Omega}^{-1}$, feasible GLS estimators of the coefficients of the deterministic components in the original model (parameters $\mu_0$, $\mu_1$ and $\theta$ of eq. (III.1)) are obtained pre-multiplying eq. (III.1) by $\tilde{Q}'\tilde{A}(L)$:

$$\tilde{Q}'\tilde{A}(L)X_t = \tilde{H}_{0t}\mu_0 + \tilde{H}_{1t}\mu_1 + \tilde{K}_t\theta + \eta_t, \tag{III.4}$$

where $\tilde{H}_{it} = \tilde{Q}\tilde{A}(L)a_{it}$ $(i = 0, 1)$ with $a_{0t} = 1$ for $t \geq 1$ and zero otherwise, and $a_{1t} = t$ for $t \geq 1$ zero otherwise. Then, feasible GLS estimators of deterministic components' coefficients are obtained (under the null of $r = r_0$) by multivariate LS applied to eq. (III.4). Once this is done, the unobserved (and uncontaminated) vector $Y_t$ can be estimated from eq. (III.1).

The authors show that the asymptotic distribution of the traditional LR test applied to $\hat{Y}_t$ is not affected by the inclusion of impulse or step dummies in the original model. This feature constitutes an important advantage over traditional procedures (see Johansen et al. (2000)) since for the latter, specific asymptotic tables need to be generated in each case as critical values depend on the break dates. This is specially undesirable for applied work since new tables are needed whenever new data points become available.

This procedure is extended by Lütkepohl et al. (2004) for the case of a unique level shift at an unknown date. The main difference with respect to Saikkonen and Lütkepohl (2000) is that the extended procedure includes an initial step in which the break date is estimated based on a VAR in the levels of the variables.

Once the break date has been estimated, Lütkepohl et al. (2004) proposal is to apply the same feasible GLS procedure as Saikkonen and Lütkepohl (2000) for determining the cointegration rank. The authors derive its asymptotic distribution and show that it is the same as the one derived by Saikkonen and Lütkepohl (2000) for the case of a known break date, which was in

turn the same as the obtained by Saikkonen and Lutkepohl (2002) for the case of no breaks.

#### III.2.1.2 Impulse Indicator Saturation

Recent developments on automatic model selection procedures applied to fully saturated regressions with impulse indicators seem to provide a general robust method to identify outlying observations that do not suffer from the drawbacks of traditional sequential searches. To determine where atypical observations may situate, the impulse indicator saturation (IIS) methodology requires the inclusion of $T$ indicator variables $d_{j,t} = 1_{\{j=t\}}$ for $j = 1, .., T$ (one indicator for each observation), in the regression model. Since a perfect fit would turn out in such a model, the indicators must be included in groups.

As described, *inter alia*, by Santos et al. (2008), in the first step only half of the indicators are included ($d_{j,t} = 1_{\{j=t\}}$ for $j = 1, .., T/2$), and those that are statistically significant at a predetermined significance level $\varphi$ ($|t_j| < c_\varphi$) are recorded. Next, the first $T/2$ indicators are dropped and those for the remaining observations are included. Finally, the significant indicators in each step are included altogether and those which are non-significant dropped.

Johansen and Nielsen (2009) show that the efficiency loss due to testing the significance of $T$ indicators is almost nonexistent for low nominal sizes of the tests ($\varphi \leq 1/T$). Indeed, in the case of no outliers and with $\varphi = 1/T$ the procedure will, on average, retain only one indicator. This has the negligible negative effect of dropping just one non-outlying observation.

Along the lines of Johansen and Nielsen (2009), Castle et al. (2012) study the characteristics of IIS but in a framework in which the other regressors are also selected with a general to specific methodology. The authors use the automatic model selection algorithm *Autometrics* which, starting from an initial General Unrestricted Model (GUM), reduces it up to a valid simpler expression (see Doornik (2009)). Their results show that non-relevant efficiency loses are caused by the inclusion of $T$ impulse indicators when they are irrelevant. Additionally, Castle et al. (2012) study the performance of *Autometrics* with IIS under the presence of outliers and find a good performance in jointly selecting variables and detecting breaks.

### III.2.2 Our empirical strategy for dealing with outlying observations in the pairwise approach

As aforementioned, the problem we are dealing with requires cointegration tests that consider the possibility of multiple outliers and breaks. In this section we propose an empirical strategy

and check its properties by Monte Carlo.

Our proposal is to jointly select the dynamic structure and the interventions applying *Autometrics* with IIS in single equation models for the first differences of all the components (the *DGUM* in Castle et al. (2012))[1]. Next, after the interventions are identified, three alternative procedures may be applied for the pairwise cointegration tests:

i) Estimate bi-variate VEqCM models including the interventions found for both variables, drop the insignificant and test for cointegration with the Johansen test. This procedure may require simulating critical values for each test. ii) Test for cointegration in a single-equation framework using the *PcGive* approach (see Kiviet and Phillips (1992)). The outliers search could be done on these equations. This does not require the simulation of new critical values, but requires exogenous variables for it is a single equation procedure. iii) Use the estimated dates and apply the GLS procedure proposed by Lütkepohl et al. (2004).

We disregard alternatives i and ii. The former due to the complexity of simulating new critical values for each test, the latter due to the absence of guarantees about the exogeneity of the variables. Then, we focus on the third alternative. In appendix B.1 we discuss a problem that this strategy may suffer under the presence of consecutive outliers and propose a strategy for dealing with it.

### III.2.3 Simulation results for the Pairwise strategy with outlier correction

To analyze the performance of the strategy proposed in §III.2.2 we hash up the simulation exercise of chapter II but applying the outlier correction strategy described above. In *DGP 1* of §II.3 we simulated a 100 dimensional VAR model (see eq. (III.5) below) in which all the series are I(1) and a subset of $n_1$ series share a single stochastic trend. The notation $n_1$ will be used both to denote the number of components in the *fully cointegrated subset* and as its label.

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Phi_1 \Delta Y_{t-1} + \epsilon_t, \tag{III.5}$$

where $\alpha$ and $\beta$ are $N \times r$ matrices; $0 < r < N$; $r$ is the number of cointegration relationships; and $\epsilon_t$ is a Gaussian white noise with covariance matrix equal to the identity. The specific forms

[1] This strategy implies the assumption that series are at most I(1).

of the matrices are:

$$\beta' = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & & \\ -1 & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}_{r \times N} \quad ; \quad \alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r}$$

and the values $\alpha_i$ are taken from the uniform distribution with parameters $[0.15, 0.3]$ — these parameters were motivated by results in Espasa and Mayo-Burgos (2013) for CPI series. Matrix $\Phi$ is diagonal with its elements taken from the uniform distribution with parameters $[0.5, 0.8]$. See chapter II for a discussion about the properties of this DGP.

In chapter II we considered four different possibilities for $n_1$; in scenario 1 $n_1 = 10$; in scenario 2, $n_1 = 15$; in scenario 3, $n_1 = 25$; and in scenario 4, $n_1 = 40$. We also considered three different sample sizes; $T = 100$, $T = 200$ and $T = 400$. As will become clear later, the experiments are too complex to consider many different alternatives, thus we do not reproduce the experiments for all scenarios and sample sizes. We focus just on scenarios 1 and 3 ($n_1 = 10$ and $n_1 = 25$ respectively), and $T = 200$. In this section we maintain the name of the scenarios (1 and 3) just to keep track with chapter II.

Furthermore, with the aim of not adding more complexity to the experiments we apply IIS without Autometrics — what Castle et al. (2011) call the 1-cut approach (see §III.2.1.2) — and check its properties by Monte Carlo.

#### III.2.3.1 Design of the experiments

To study the behavior of the pairwise strategy under the presence of outliers we contaminate the series used in chapter II with 1, 2 or 3 level shifts of size $\gamma = 4\sigma$ or $\gamma = 5\sigma$, with $\sigma$ being the standard error for the univariate model of the uncontaminated series. For doing this we proceed as follows: i) Apply 1-cut IIS with $\varphi = 1/T$ to the first differences of the 164 basic components of the US CPI and store the percentage of series that have at least 1, 2 or 3 level breaks. The results were $[87\%, 65\%, 42\%]$ respectively. ii) Take the simulated series $(y_{i,t})$ of

chapter II and contaminate them using the corresponding percentage as:

$$x_{i,t} = y_{i,t} + \Gamma LS_{t^*},$$

with $\Gamma$ and $LS_{t^*}$ being $(1 \times q)$ and $(q \times 1)$ vectors respectively; where $q$ takes the values 1, 2 or 3. The vectors in $LS_{t^*}$ take the value zero up to the date of the break and one from then on. The dates of the breaks ($t^*$) are randomly set in the interval $[20, 180]$, but this random position is maintained for all the 1000 replicas. iii) The percentage of contaminated series is maintained for series inside and outside $n_1$.

With the contaminated series we replicate the simulations of chapter II in four different ways: a) *True Dates*: the true outliers' position is used to apply the pairwise strategy with outlier treatment. b) *Estim Dates*: outliers' positions are estimated using *1-cut* IIS in single equation models with three lags and the pairwise procedure with outliers' treatment is applied (in appendix B.2 we study the performance of the *1-cut* approach applied to our simulated series). c) *No Outl:* the pairwise procedure with outlier treatment (using estimated dates in b)) is applied to the uncontaminated series. d) *No Corr*: the original pairwise procedure (with no outlier treatment) is applied to the contaminated series.

The number of pairs among the 100 series in the DGP is 4950, and for each of the 1000 replicas, we need to apply the GLS procedure described in §III.2.1.1 to all the pairs. Considering the 1000 replicas and the three cases a) to c), the number of times that we need to apply the GLS procedure is $1000 \times 4950 \times 3 = 14.85$ *million*. To this figure we need to add case d) in which the GLS procedure is not applied but still we need to perform $1000 \times 4950 = 4.95$ *million* cointegration tests. In all cases, the lag length is unknown and we estimate it with the AIC admitting a maximum of five lags, what multiplies the number of estimations by five. This complexity of the simulation experiments led us not to consider other possible DGPs or sample sizes.

#### III.2.3.2 Results

We assess the performance of the procedure using what Castle et al. (2011) call *potency* and *gauge* in the context of model selection. While *gauge* measures the retention frequency of irrelevant variables when selecting among a — potentially large — set of candidates, *potency* denotes the average retention frequency of relevant variables. In our context *gauge* measures

the inclusion of wrong components in the estimated *fully cointegrated subset* and *potency* the inclusion of correct components.

Table III.1 includes the *gauge* and *potency* of the pairwise procedure with outlier treatment. Five main conclusions emerge from the table: i) Using the outlier correction treatment when it is not required (block *Number of breaks: 0* of the table), somewhat deteriorates the procedure's performance with respect to not correcting for outliers (the case of chapter II). While when correcting for outliers we include on average $0.917 \times 10 = 9.2$ and $0.885 \times 25 = 22.1$ correct series in scenarios 1 and 3 respectively, if no correction is used these figures increase to $0.969 \times 10 = 9.7$ and $0.956 \times 25 = 23.9$ (see table II.3 in chapter II). ii) When using the *True Dates* with only one break per series the procedure's performance is similar to the case with no breaks (compare with table II.3 in chapter II). iii) Though *gauge* remains at very low levels, *potency* deteriorates with the number of outliers, even when using the *True Dates*. iv) When using the *Estim Dates* — instead of the *True Dates* — *potency* reductions are observed. But recall that, since we are using the *1-cut* approach, we are not selecting the lag length for the series in $n_1$ — which are the most relevant in determining the pairwise procedure's potency — so that potencies under *Estim Dates* columns are just lower bounds for those that will be obtained when using *Autometrics*. This is more relevant for scenario 3, as the proportion of series in $n_1$ is larger. v) The comparison between columns under *Estim Dates* and those under *No Corr* shows that, although we have only lower bound potencies for *Estim Dates*, estimating the dates is better than nothing. This conclusion is not true when the number of breaks is low (see block *Number of breaks: 1*). Additionally, we find that *No Corr* slightly outperforms *Estim Dates* for two breaks and $\gamma = 4$ in scenario 3. This last result reinforces the argument that not selecting the dynamic structure is more damaging for scenario 3, for which the proportion of series in $n_1$ is larger.

To conclude, in this section we proposed an outlier treatment for the pairwise strategy and studied its properties by Monte Carlo. Though the proposal requires the identification of the outliers' dates by *Autometrics* with IIS, the use of *Autometrics* in our simulation setting can be very tedious. As an alternative, we applied IIS with the simple 1-cut approach. The main disadvantage of this procedure is that it does not jointly select the dynamic structure and the impulses. In series with richer dynamics (those in $n_1$) this problem seems to be more important, as our results show. For this reason we interpret potencies in table III.1 as lower bounds for those that will be obtanied with *Autometrics*. As a general conclusion from table III.1 we can say

Table III.1: Gauge and Potency of the pairwise procedure with outlier treatment

| | **Scenario 1**, ($n_1 = 10$) | | | | | | **Scenario 3**, ($n_1 = 25$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **True dates** | | **Estim dates** | | **No corr** | | **True dates** | | **Estim dates** | | **No corr** | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Number of breaks: 0 | | | 0.3 | 91.7 | | | | | 0.2 | 88.5 | | |
| Number of breaks: 1 | | | | | | | | | | | | |
| $\gamma$ =4 | 0.3 | 95.4 | 0.3 | 85.6 | 0.3 | 92.4 | 0.2 | 93.2 | 0.2 | 78.5 | 0.2 | 88.1 |
| $\gamma$ =5 | 0.3 | 95.3 | 0.3 | 85.0 | 0.3 | 88.3 | 0.2 | 93.1 | 0.2 | 78.3 | 0.2 | 82.3 |
| Number of breaks: 2 | | | | | | | | | | | | |
| $\gamma$ =4 | 0.3 | 91.2 | 0.3 | 77.0 | 0.3 | 75.6 | 0.2 | 88.6 | 0.2 | 70.2 | 0.2 | 72.9 |
| $\gamma$ =5 | 0.3 | 90.4 | 0.4 | 77.2 | 0.4 | 64.5 | 0.2 | 87.8 | 0.2 | 70.6 | 0.2 | 63.0 |
| Number of breaks: 3 | | | | | | | | | | | | |
| $\gamma$ =4 | 0.3 | 87.5 | 0.4 | 74.5 | 0.4 | 72.3 | 0.2 | 82.7 | 0.2 | 69.1 | 0.2 | 68.5 |
| $\gamma$ =5 | 0.3 | 86.2 | 0.3 | 78.6 | 0.4 | 66.7 | 0.2 | 81.4 | 0.2 | 73.1 | 0.2 | 64.2 |

- *Pot* = $\frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$, with $Z_{1,i}$ being the number of correct series included in $\hat{n}_1$ (the estimated fully cointegrated subset) in experiment *i*.
- *Gauge* = $\frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$, with $Z_{2,i}$ being the number of incorrect series included in $\hat{n}_1$ in experiment *i*.
- *True Dates*: the true outliers' position is used to apply the pairwise strategy with outlier treatment.
- *Estim Dates*: outliers' positions are estimated using 1-cut IIS.
- *No Corr*: the original pairwise procedure (with no outlier treatment) is applied to the contaminated series.
- Figures under columns *Estim Dates* have to be interpreted as a worst possible case because we are not selecting the dynamic structure for series in $n_1$, what deteriorates IIS's potency. Note that series in $n_1$ are in fact the most important series to determine the potency of the pairwise strategy.

that the presence of outliers distorts the potency of the pairwise procedure and this distortion is larger, the larger the number of outliers. Some procedure is required for diminishing this problem when series are contaminated. Our results show that the proposal sketched in §III.2.2 can be regarded as an acceptable approach.

## III.3 Small samples correction

As noted in chapter II, the potency of the pairwise procedure to discover the true *fully cointegrated subset(s)* deteriorates as $T$ decreases. There are two reasons that explain this issue: first, the Johansen's trace test properties deteriorates in small samples; second, the equivalence of the tests showed in chapter II (see theorem 1) and summarized in the introduction of this chapter is valid only asymptotically, and also deteriorates in small samples.

To mitigate this problem, we propose a slight modification of the procedure: relax the '*full*

*cointegration*' requirement to '*almost full cointegration*'. The relaxation consists of allowing to enter in $\hat{n}_1$ those series for which cointegration with at most $\lambda$ series in the initially estimated fully cointegrated subset was not found at the original $\varphi$ nominal size, but it is found if test are performed at $\varphi^*$ ($\varphi^* > \varphi$) — a detailed description of the relaxation procedure is included in §III.5. This strategy will lead to increase the potency of the procedure but will also increase the risk of including wrong series.

Assume that we have three $I(1)$ series, $S_1, S_2$ and $S_3$, such that $S_1, S_2$ is the unique truly cointegrated pair. Assume further that we wrongly find cointegration for the pair $S_1, S_3$. As we — wrongly — found the stochastic trend of $S_3$ to be that of $S_1$, which is also the same as that of $S_2$, the probability of finding cointegration for the pair $S_2, S_3$ — given that we found cointegration in $S_1, S_3$ — would be larger than or equal to $\varphi$ (the unconditional asymptotic probability). Call this conditional probability $\tilde{\varphi}$.

Let $Z_2^\lambda$ be the number of series that do not belong to the true fully cointegrated subset but cointegration tests indicate cointegration with all but $\lambda$ of the series in the original $\hat{n}_1$. That is, $Z_2^\lambda$ represents the number of potential candidates to enter the *almost fully cointegrated subset* that we do not want to include. For a given initial $\hat{n}_1$, the larger $\tilde{\varphi}$ is, the larger $Z_2^\lambda$ would be. Similarly, for a given $\tilde{\varphi}$, the smaller $\hat{n}_1$ is, the larger $Z_2^\lambda$ would be, hence, $Z_2^\lambda$ is a decreasing function of the original $\hat{n}_1$.

Table III.2 illustrates these arguments. It shows the mean number of potential candidates for each of the four scenarios considered in chapter II for *DGP 1* and relaxation parameter up to $\lambda = 3$ (see §III.2 for a brief description of the DGP, or chapter II for a more detailed discussion). We consider only the sample size $T = 100$ because for larger samples in chapter II we did not found relevant potency loses. While column (a) of the table contains the mean number of series that have between 1 and $\lambda$ holes in the current *(almost) fully cointegrated set*, column (b) includes the series of column (a) whose holes were filled after relaxing cointegration tests from 1% to 5% of significance. Columns (c) and (d) are analog to (a) and (b) but wrong candidates are excluded.

As the *Ratios* columns show, while for scenarios 3 and 4 (large $n_1$) almost all the potential candidates are correct series, this is not true for scenarios 1 and 2. The difference between scenarios becomes more evident for larger relaxation parameters ($\lambda$). For instance, with $\lambda = 1$, in scenario 1, 75% of the candidates are correct series, whereas in scenario 4, 98% of the

candidates are correct ones. For $\lambda = 3$, while in scenario 1 only 40% of the candidates are correct series, in scenario 4, 97% of them are so.

Table III.2: Statistics of the Relaxation process. Mean number of potential candidates ($T = 100$)

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 1$**

| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.05 | (c) No Rest | (d) pval>0.05 | c/a | d/b |
| Sce 1 | 1.06 | 0.92 | 0.76 | 0.68 | 0.75 | 0.75 |
| Sce 2 | 1.37 | 1.26 | 1.18 | 1.10 | 0.87 | 0.88 |
| Sce 3 | 1.96 | 1.83 | 1.85 | 1.72 | 0.95 | 0.95 |
| Sce 4 | 2.63 | 2.51 | 2.58 | 2.46 | 0.98 | 0.98 |

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 2$**

| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.05 | (c) No Rest | (d) pval>0.05 | c/a | d/b |
| Sce 1 | 1.24 | 0.76 | 0.73 | 0.51 | 0.65 | 0.70 |
| Sce 2 | 1.60 | 1.17 | 1.26 | 0.98 | 0.82 | 0.85 |
| Sce 3 | 2.26 | 1.80 | 2.08 | 1.70 | 0.94 | 0.94 |
| Sce 4 | 3.30 | 2.77 | 3.19 | 2.69 | 0.97 | 0.97 |

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 3$**

| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.05 | (c) No Rest | (d) pval>0.05 | c/a | d/b |
| Sce 1 | 0.85 | 0.26 | 0.29 | 0.10 | 0.42 | 0.40 |
| Sce 2 | 1.27 | 0.56 | 0.86 | 0.43 | 0.72 | 0.75 |
| Sce 3 | 1.75 | 1.00 | 1.57 | 0.91 | 0.91 | 0.91 |
| Sce 4 | 2.48 | 1.60 | 2.39 | 1.56 | 0.96 | 0.97 |

- Number of experiments: 1000. For scenarios 1 to 4, $n_1$ is $10, 15, 25$ and 40 respectively, with $n_1$ being the number of series that share a single common trend among the 100 series in the model.
- The $\hat{n}_1$ subset is updated in each step.
- Column (a) contains the mean number of series that have between 1 and $\lambda$ holes in the current *(almost) fully cointegrated set* (cointegration at the 1% was rejected with at least $\lambda$ series in $\hat{n}_1$).
- Series in column (b) and (d) are those of column (a) and (d) whose holes were 'filled' after relaxing cointegration tests to the 5%.
- A comparison between columns (a) and (b) or columns (c) and (d) gives an idea of the effects of requiring cointegration at the 5% for the holes to be filled vs. no requiring anything.
- Columns (c) and (d) are analog to (a) and (b) but only truly correct series are considered.
- Note that $Z_2^\lambda$ does not explicitly appears in this table, it can be obtained by subtracting column (c) to column (a), or column (d) to column (d).
- All figures (including the *Ratios*) are averages across experiments. Then, figures in column *Ratios* are not necessarily equal to $c/a$ and $d/b$ because they are the mean across experiments.

Hence, results in table III.2 confirm that it is for situations with 'large' initial $\hat{n}_1$ that the relaxation is less risky, i.e, $Z_2^\lambda$ is a decreasing function of $n_1$. Notably, it also happens that it is precisely for those scenarios that improving the original results is most needed. Recall that the equivalence of cointegration tests studied in theorem 1 of chapter II is valid only asymptotically.

Then, as $T$ decreases, the asymptotic equivalence deteriorates and the probability of finding cointegration between all the correct pairs moves away from $(1-\varphi)$ and becomes a function of the number of pairs. The larger the number of pairs, the lower the probability of finding cointegration between all (or a high proportion) of them.

In line with this argument, results in table II.3 of chapter II show that the original potencies for $T = 100$ are decreasing in $n_1$: $0.77, 0.73, 0.67$ and $0.62$ for scenarios 1 to 4, respectively.

Table III.3 adds more evidence for the two arguments made above, namely, while the risk of relaxing the full cointegration requirement is decreasing in $n_1$, the potential benefit is increasing. Define $Z_1$ as the number of correct series included in $\hat{n}_1$, and $Z_2$ as the number of wrong series included in that subset. The table contains the ratios $Z_2/n_1$ and $Z_1/n_1$ as a function of $\lambda$ for the four scenarios. As it shows, in scenarios 3 and 4 we can increase the ratio $Z_1/n_1$ (potency) by 20 percentage points with almost no cost in terms of $Z_2/n_1$. This is not the case for scenarios 1 and 2, for which the benefits are lower and the costs somewhat higher.

Table III.3: Mean of the ratios $Z_1/n_1$ and $Z_2/n_1$ as a function of the relaxation parameter $\lambda$

| | **Mean $Z_2/n_1$** | | | | **Mean $Z_1/n_1$** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Sce1** | **Sce2** | **Sce3** | **Sce4** | **Sce1** | **Sce2** | **Sce3** | **Sce4** |
| $\lambda = 0$ | 0.04 | 0.02 | 0.01 | 0.00 | 0.77 | 0.73 | 0.67 | 0.62 |
| $\lambda = 1$ | 0.06 | 0.02 | 0.01 | 0.00 | 0.83 | 0.79 | 0.73 | 0.68 |
| $\lambda = 2$ | 0.08 | 0.03 | 0.01 | 0.00 | 0.87 | 0.84 | 0.78 | 0.73 |
| $\lambda = 3$ | — | 0.04 | 0.01 | 0.01 | — | 0.86 | 0.81 | 0.76 |
| $\lambda = 4$ | — | 0.05 | 0.02 | 0.01 | — | 0.87 | 0.83 | 0.78 |
| $\lambda = 5$ | — | 0.05 | 0.02 | 0.01 | — | 0.87 | 0.84 | 0.80 |
| $\lambda = 6$ | — | — | 0.02 | 0.01 | — | — | 0.85 | 0.81 |

$Z_2$ = number of wrong series included in $\hat{n}_1$
$Z_1$ = number of correct series included in $\hat{n}_1$

Therefore, the relaxation parameter $\lambda$ (which indicates the maximum number of 'holes' that a candidate series can have to enter the *almost fully cointegrated* subset) has to be defined as a function of the original $n_1$. However, as we have no prior rules to define that function, we perform a simulation exercise to decide on the appropriate $\lambda$ given the initial $\hat{n}_1$ (for the true $n_1$ is unknown in empirical applications).

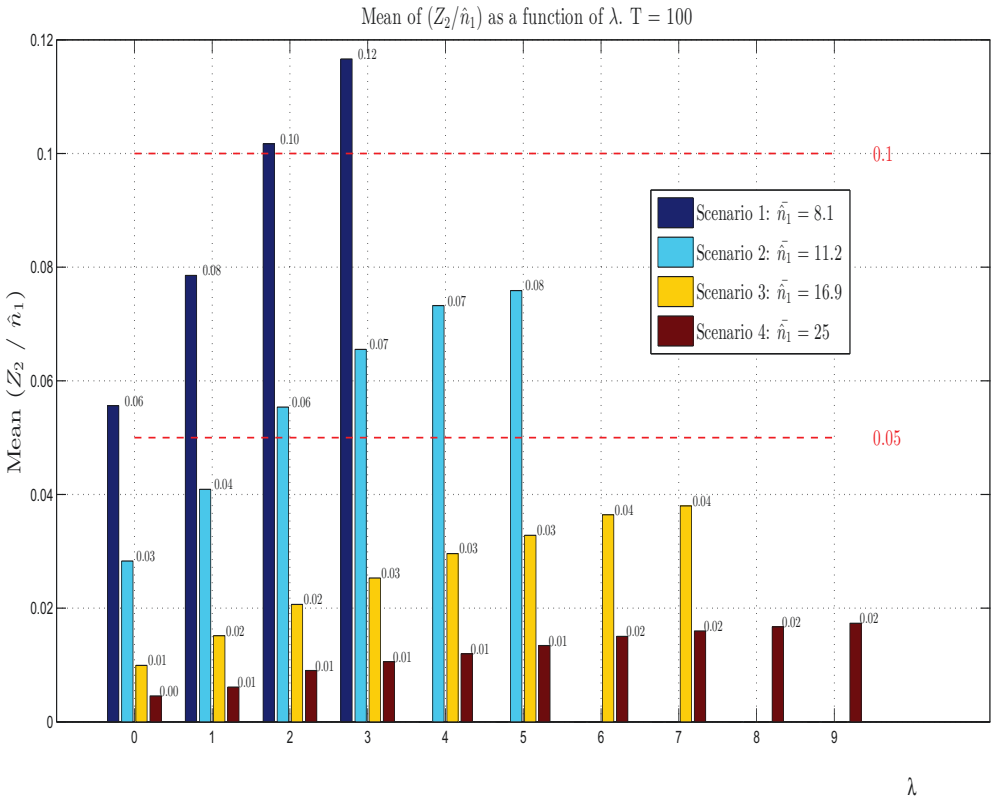
Using the same simulated series as those of chapter II we run the pairwise procedure but instead of requiring *full cointegration* we consider the relaxation to *almost full cointegration* using alternative relaxation parameters $\lambda$. The alternative values of $\lambda$ considered were; $[1, 2, ..., 9]$.

To decide on the 'optimal' $\lambda$ we consider the following criterion. For each experiment we count the number of incorrect series ($Z_2$) included in the *estimated almost full cointegrated subset* and compute the ratio ($Z_2/\hat{n}_1$), where $\hat{n}_1$ is the size of the originally estimated *strict* full cointegrated subset[2]. Next, we average this ratio over all the 1000 experiments. Finally, a decision rule to choose the optimal $\lambda$ has to be defined.

We define the optimal $\lambda$ as the maximum one such that the mean ratio $Z_2/\hat{n}_1$ does not exceed a certain threshold. Figure III.1 includes the simulation results for the four scenarios and $T = 100$. Dashed red lines represent two arbitrary decision rules to choose the optimal $\lambda$ given the initial $\hat{n}_1$. We are requiring the expected value for the ratio $Z_2/\hat{n}_1$ to be 0.05 (0.1).

With the criterion of 0.05 the optimal $\lambda$ for scenarios 1 and 2 would be 0 and 1 respectively. However for scenarios 3 and 4 this rule is not operative since we never reach the 0.05 threshold. In these cases we set $\lambda$ equal to 5 and 7 respectively since these are the $\lambda's$ for which the ratio $Z_1/\hat{n}_1$ stabilizes (see also table III.3). This sort of 'jump' in the maximum number of holes admitted (from zero and one in scenarios 1 and 2, to five and seven in scenarios 3 and 4) confirms once again the argument discussed above; the probability of having a wrong series with few holes is a decreasing function of $n_1$.

[2]Another alternative could be to compute $Z_2/n_1$, but as in practice the only possible baseline is $\hat{n}_1$ we prefer the previous criteria which turns out to be more conservative since we found $\hat{n}_1 < n_1$ in all experiments.




Note: Dashed red lines represent two arbitrary decision rules to choose the optimal $\lambda$ given the initial $\hat{n}_1$. We are requiring the expected value for the ratio $Z_2/\hat{n}_1$ to be 0.05 (0.1).

Figure III.1: Mean of the ratio (Num of incorrect series / Num of series in $\hat{n}_1$ in the strict full cointegration framework) as a function of the relaxation parameter $\lambda$.

## III.4 Generalization of the cointegration tests

This section has two objectives; propose some extensions to the pairwise cointegration tests that could be useful for empirical applications, and define some criteria to evaluate the 'quality' cointegration relationships found when working with real data.

### III.4.1 Extensions of the pairwise tests

We extend the original procedure in three directions motivated by results in Aron and Muellbauer (2013). First, we consider the inclusion of a weakly exogenous variable in all the bi-variate models, second, we allow for a *parsimonious long lag* parametrization ($PLL$), and finally we consider the possibility of a specific form of non-linearity.

#### III.4.1.1 Weakly exogenous variable

A third variable $EXO_t$ in the originally bi-variate VARs may help to find cointegration between two components of a macro variable. In models where $EXO_t$ is significant in the 'long-run' we

assess its weak exogeneity. If weak exogeneity is rejected we do not consider it in the model[3].

Note that now we may have zero, one or two common trends between two components and $EXO_t$. When there is only one, we are in the regular case and blocks of components can be constructed as summarized in §III.5. Since we are considering only the cases where $EXO_t$ is weakly exogenous, the common trend will be generated by the accumulation of the shocks to this variable.

For the case of two common trends, blocks can still be constructed and components inside them will share the two trends; one generated by the shocks of $EXO_t$ and the other by a combination of the shocks in the two components (see appendix B.3 for a detailed discussion about the inclusion of weakly exogenous variables).

Therefore, when including a third (weakly exogenous) variable the strategy for constructing the blocks of components is slightly changed. We first consider the models that have only one common trend and then those with two common trends.

Finally, note that the inclusion of $EXO_t$ can be implemented in two alternative fashions. The simplest one is just adding $EXO_t$ to all the pairs. Note however that this strategy could distort the results for the pairs that do not need this third variable. Hence, the second alternative is to proceed sequentially in two steps; first, the block search is carried out not including $EXO_t$ and then, using only the series not belonging to any (almost) fully cointegrated block, the procedure is repeated but including $EXO_t$.

#### III.4.1.2 Parsimonious Longer Lags (PLL)

As argued by Aron and Muellbauer (2013) a possible way of tackling the 'curse of dimensionality' present in VAR models with long lags is to impose 'parsimonious longer lag' ($PLL$) restrictions. That is, imposing equality restrictions on the coefficients' matrices associated with some lags. For example, instead of including all the lags from the $m$ to the $m+k$, we could include the variable $\Delta_k X_{t-m}$. This will restrict the coefficients of $\Delta X_{t-m}$ to $\Delta X_{t-m-k}$ to be the same. Of course, we need to decide which $k$ and $m$ to use. One possibility is to try with all the possible restrictions between consecutive lags and pick the best one. This option will be computationally costly in the pairwise approach framework.

Instead, we proceed as Aron and Muellbauer (2013), allowing full generality for short lags

[3] $EXO_t$ is also considered for the outlier correction procedure.

(and the seasonal) and restricting all the others. For the US CPI Aron and Muellbauer (2013) admit full generality at lags one, two and twelve; replace lags three to five with $\Delta_2 X_{t-3}$; replace lags six to eleven with $\Delta_5 X_{t-6}$; and finally, replaces lags thirteen to twenty five with $\Delta_{12} X_{t-13}$. Compared to unrestricted lags up to 25 months, 25 coefficient matrices are replaced by six.

This extension raises the problem highlighted by Nielsen and Nielsen (2008) about the appearance of large, albeit not significant, roots in the characteristic polynomial of the VAR. This issue is particularly relevant for this application since, as we propose in §III.4.2, we will disregard models with large second (third) roots. We tackle this issue by computing the roots in the polynomial that remains after disregarding the $PLL$ structure. This procedure assumes that the polynomial associated with the $PLL$ does not have unit roots.

#### Other possible restrictions in the coefficients

The PLL structure suggested by Aron and Muellbauer (2013) is an interesting way of restricting the coefficients, but it is just one possibility among an undefined number of alternatives. A more general way of considering PLL structures could be to try all possibilities for restricting consecutive pairs, triplets, etc, and select the best model.

Yet, another possibility is to consider moving average structures, by restricting the coefficients of consecutive lags to be exponentially decreasing. Assuming that the absolute value of the MA coefficient ($\theta$) is between 0.4 and 0.8, we could try $\theta = [-0.8, -0.7, ..., -0.4, 0.4, ..., 0.8]$ estimate the restricted AR model in each case and keep the best model.

In order to not increase the complexity of our empirical applications, we do not explore these possibilities.

### III.4.1.3 Non-linearities

As argued by Aron and Muellbauer (2013), when dealing with prices, the possibility that producers adjust prices more frequently when cost changes are more volatile (see also Reis (2006)) may induce non-linearities in inflation; high recent inflation would induce disproportionately high future inflation. A simple way of considering this possibility is the one suggested by Aron and Muellbauer (2013); regress $(\Delta_6 log(P))^2$ on a constant and $\Delta_6 log(P)$, where $P$ is a vector of two disaggregated prices. The residual of this regression and its 6-month lag may capture the non-linearities described above (see Aron and Muellbauer (2013) for further details) and can be included as purely exogenous variables in the bi-variate VARs.

### III.4.2 Quality assessment of the cointegration tests' results

In order to obtain economically and statistically sensible cointegration restrictions we propose to consider only those that satisfy the following conditions: i) the 'long-run' relationship does not require a deterministic trend; ii) coefficients of both components are statistically significant; iii) the second largest root of the bivariate VAR's characteristic polynomial is not close to one; and iv) the 'long-run' relationship is stable over time.

The first condition is important when dealing with macroeconomic variables as the most general case is that they show systematic growth, so some procedure for dealing with deterministic terms should be considered. When dealing with prices (as it is the case of the application in this chapter), the inclusion of a linear trend in the cointegration relationship should be considered with extreme caution, since forecasts will show a price systematically increasing over the other. Unless there are strong theoretical foundations for such a forecast, our suggestion is not to consider cointegration relationships including linear trends. Then, we proceed as follows: a) Estimate all pairwise VEqCM models (under $r = 1$) including a trend in the cointegrating space. b) Test the significance of the trend and disregard that pair as being 'purely' cointegrated if the trend is required. c) For the pairs that do not require a trend, test for cointegration not including the trend.

This strategy is not exactly the one suggested by Nielsen and Rahbek (2000). These authors find that cointegration rank tests are asymptotically similar with respect to the parameters of deterministic components. Thus, they proceed in two steps. First, they test the cointegration rank in a model that includes all the deterministic components (constant, trends and interventions) in the cointegrating relationships and its differences in the VAR. Second, once the cointegration rank is determined, hypothesis on deterministic parameters can be tested (see also Doornik et al. (1998) and Juselius (2006)). The reason for not exactly following this procedure is that we are interested in testing cointegration only if the model does not require a trend in the cointegrating space.

With condition ii we want to exclude stationary variables from the *fully cointegrated subsets*. Tests on the coefficients can be performed as in Johansen (1995) at, say, 10% of significance not to exclude too many pairs.

The third condition is relevant for the procedure does not exclude a priori series with an $I(2)$ behavior. Our proposal is to disregard models whose second root is larger than, say, 0.90.

For condition iv (stability of the 'long-run' relationships), we propose to consider both, the first eigenvalue and the stability of the long-run coefficients. Pairs that do not pass both tests should be disregarded. The reason for considering both tests is that constant eigenvalues do not imply constant long-run coefficients, and constancy of those coefficients does not imply that Johansen's test results are stable over time (see Hansen and Johansen (1999) and Juselius (2006)). Eigenvalues are assessed by the *fluctuation test* proposed by Hansen and Johansen (1999), and long-run coefficients in the evaluation period are compared with the full sample estimation (see Hansen and Johansen (1999) and Juselius (2006)). For the empirical application in §III.6, in both cases we consider forward recursive tests at 5% of significance based on the concentrated model and the evaluation period is the last five years of the sample (see Juselius (2006) for a discussion on the pros and cons of using the concentrated model *versus* the full model).

## III.5 Detailed algorithm of the Pairwise procedure

Since the original procedure has been modified in several ways to make it more useful for empirical applications, we devote this section for describing the final algorithm. Before doing so, a comment about seasonal unit roots is worthwhile.

The procedure studied in this thesis does not deal with the possibility of seasonal unit roots. A proper treatment of this issue will highly increase its complexity, specially when testing for cointegration between series with different number of seasonal unit roots. Seasonally adjusted series may artificially generate common dynamics in the series and therefore distort estimations. Nonetheless, in order to avoid complex estimation problems, those kind of series are sometimes used by econometricians. Just to cite some articles, Hendry and Hubrich (2011), Stock and Watson (2007) and Trenkler et al. (2007) are relevant examples. Other, probably better, alternative could be to seasonally adjust the series using past filters only, but this is also complex, for statistical offices do not provide these series and the standard software do not allow to do that. For these reasons our approach is to use seasonal dummies.

The procedure involves ten steps:

i. Perform Johansen cointegration tests between all possible pairs of components, disregarding those which do not pass the four quality conditions in §III.4.2, and store the resulting p-values.

ii. Construct a $N \times N$ boolean adjacency matrix, $A$, that contains a 1 if the corresponding pair is cointegrated and zero otherwise.

iii. Find the *maximal clique* on $A$ using, for example, the Bron-Kerbosch algorithm (see Bron and Kerbosch (1973)). The *maximal clique* is defined as the largest subgraph in which all nodes are pairwise connected (see also, Bollobás and Erdös (1976)). We rename the maximal clique as *fully cointegrated subset*, $\hat{n}_1$.

iv. As analyzed in §III.3, in relatively short samples it may be desirable to relax the requirement of *full cointegration* and let components that are cointegrated with *almost all* the other elements of the *fully cointegrated subset* to enter the subset. Call this new set *almost fully cointegrated*. If the user does not want to consider this relaxation, in point iii, instead of finding just the largest clique, all independent cliques should be found, and the procedure ends there. Otherwise, continue.

v. Define the relaxation parameter ($1 \leqslant \lambda < \hat{n}_1$, with $\hat{n}_1$ being the number of series in the estimated fully cointegrated subset) to identify the candidates to enter in the *almost fully cointegrated* set. A series outside the original set is a candidate if it satisfies two conditions:

   (a) cointegration — at the original $\varphi$ of confidence — is rejected with at most $\lambda$ of the series already in the subset $\hat{n}_1$.

   (b) when the nominal size of the cointegration test is relaxed to $\varphi^*$ the candidate is cointegrated with all the series already in the subset $\hat{n}_1$.

vi. Construct the set of candidates $C_0$. If all the candidates are pairwise cointegrated between each other (at the original $\varphi$), let all of them in and go to point ix (because there are not more potential candidates).

vii. If not, find the maximal clique (see point iii) inside $C_0$ and let in all the series in the maximal clique. Note that after including these series there could still remain some potential candidates, so check for this possibility; construct a new set of candidates $C_1$, and go to previous point.

viii. If there are not cointegrated candidates, try to include them sequentially starting with the one which is cointegrated with most components of series already in the set. In case of conflict (there are candidates that are cointegrated with the same number of variables

already in the set), use the p-values stored in step $i$ to decide. An adhoc criteria could be, for example, to include the series whose sum of p-values for the null $r = 0$ in cointegration tests with the series already in (or with the ones for which cointegration was rejected) is the maximum. Other adhoc possibility could be to include the series whose sum of p-values for the null $r = 1$ in cointegration tests with the series already in (or with the ones for which cointegration was rejected) is the minimum.

ix. Repeat steps iii to viii but excluding the series already included in some almost fully cointegrated set.

x. Once the disaggregation map is obtained, the forecasting equations can be constructed.

Figure III.2 summarizes the algorithm in five basic steps.

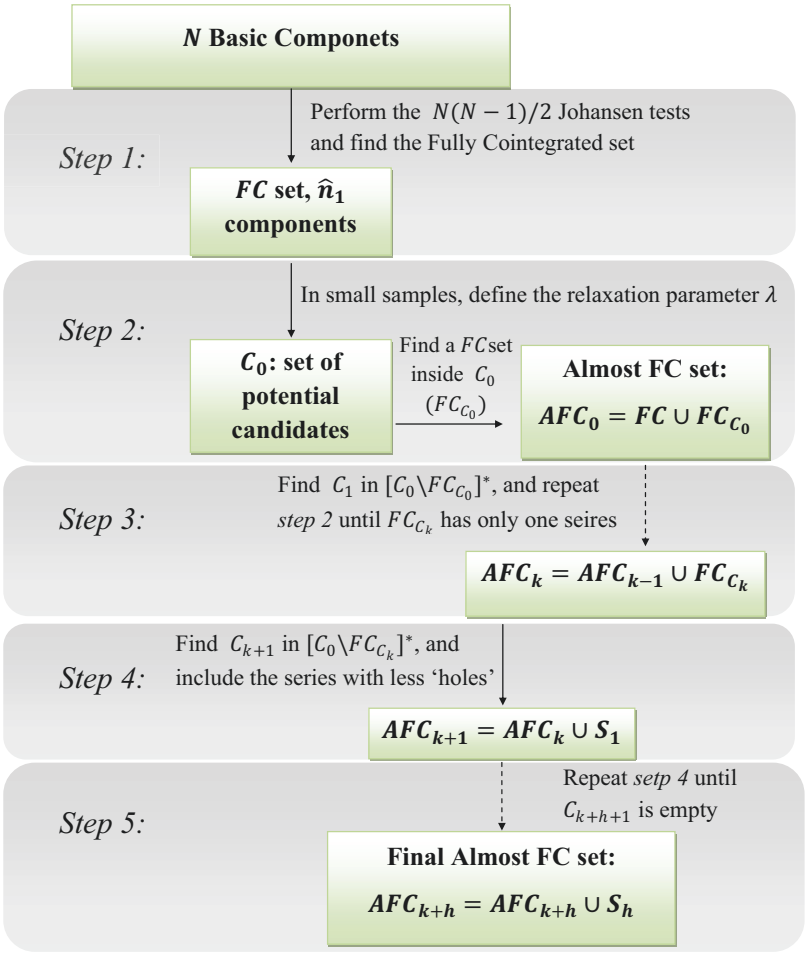



Figure III.2: Sketch diagram for the algorithm to construct the set of components with a common trend

Notes:
- * The symbol '\' represents the set difference operator, so that $A \setminus B = \{x \in A : x \notin B\}$.
- In step 1 disregard pairs which do not pass the quality tests described in §III.4.2.

## III.6 Empirical application: US CPI

In this section we apply the pairwise procedure with outliers correction to the US CPI. The absence of economic theory linking disaggregated prices in the long run could make the concept of *cointegration* to sound inadequate for this application. However, this observation does not preclude the existence of linear combinations between CPI components that cancel unit roots and could be useful to obtain better forecasting results. The absence of theory only implies that these relationships may not be expected to be *permanent* as, for example, is the relationship between income and consumption. For this reason, in this section we substitute the concept of

cointegration by *common unit roots restrictions*[4].

### III.6.1 Data

The CPI break down used in this analysis correspond to the maximum disaggregation level available to the public in the *Bureau of Labor Statistics* (seasonally un-adjusted CPI-U for all urban consumers) for the period 1999.1 – 2014.12 (192 observations). The total number of components is 181. After dropping those with less than 162 valid observations we keep 172 basic components. From these series we exclude eight that evolve by steps (regulated prices) so that we end up with 164 series which, considering 2014 weights, represent 91% of the CPI [5].

Figure III.3 describes all the 164 components and the aggregated CPI. The first panel shows a great heterogeneity among the components, what highlights the importance of disaggregated analysis.

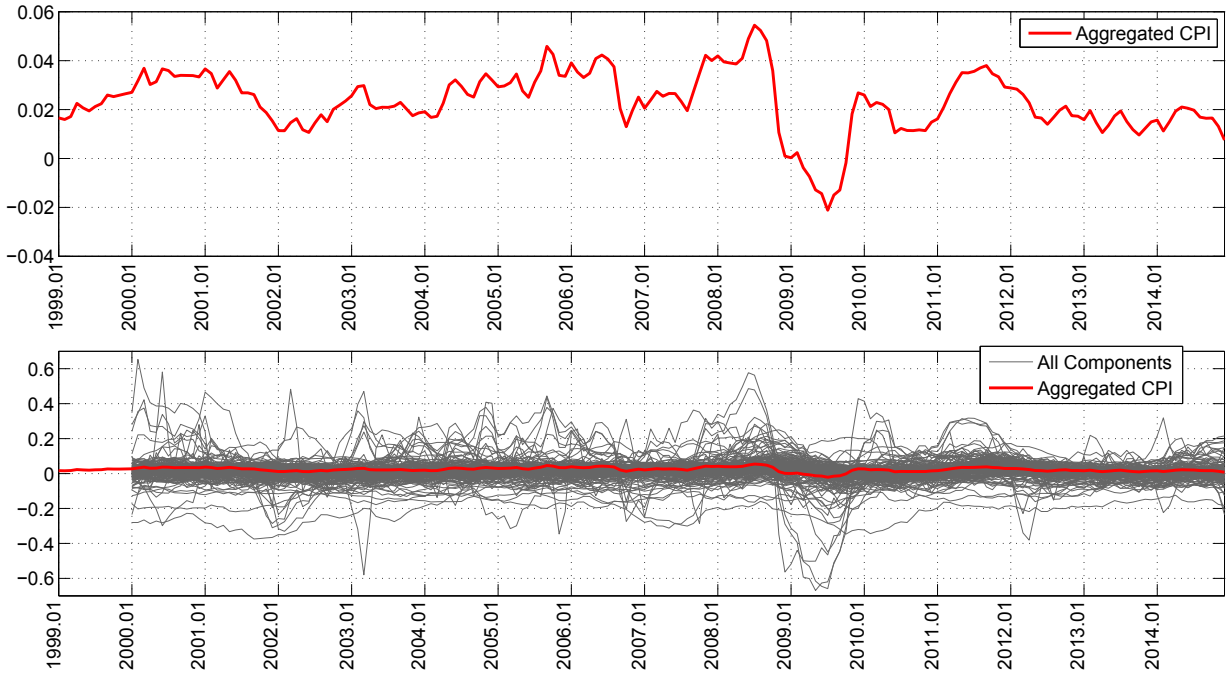



Figure III.3: $\Delta_{12}(logCPI)$ and its components, 1999.1 - 2014.12

[4]We are grateful to David Hendry for this observation.

[5]The eight excluded series are: College tuition and fees, Elementary and high school tuition and fees, Child care and nursery school, Postage, Delivery services, Wireless telephone services, Food at employee sites and schools, and Housing at school excluding board.

### III.6.2 Outliers' analysis

As described in §III.2.2 we carry out the outliers search for the 164 components in individual models for the differenced components using *Autometrics* with Impulse Indicator Saturation (IIS). We select the impulses in two steps. First, we use a target gauge of 0.5% to select lags, seasonal dummies and impulses, and store the retained impulses. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.

To make tables legible, basic components are grouped into six broad categories: non-energy industrial durable goods (MAN Dur), processed food (PF), services (SERV), non-processed food (NPF), non-energy industrial non-durable goods (MAN No Dur) and energy (ENE)[6].

Table III.4 summarizes the results. Four main observations emerge from it: i) the average number of outliers per series is 4.7 (2.6% of the observations), ii) energy and services prices are the most contaminated with a mean proportion of 4.5% and 3.6% of outlying observations per component respectively, iii) 37% of the outliers are large (larger than $4\sigma$ in absolute value), and iv) large outliers are more typical in services and energy prices representing 52% and 45% of the total number of outliers respectively.

A list of 'highly contaminated' series (5% or more outlying observations) is included in table B.2 of appendix B.4. These are 21 out of the 164 series we are dealing with and their weight represent 8.15% of the CPI.

Table III.4: Mean number of outliers by size and category

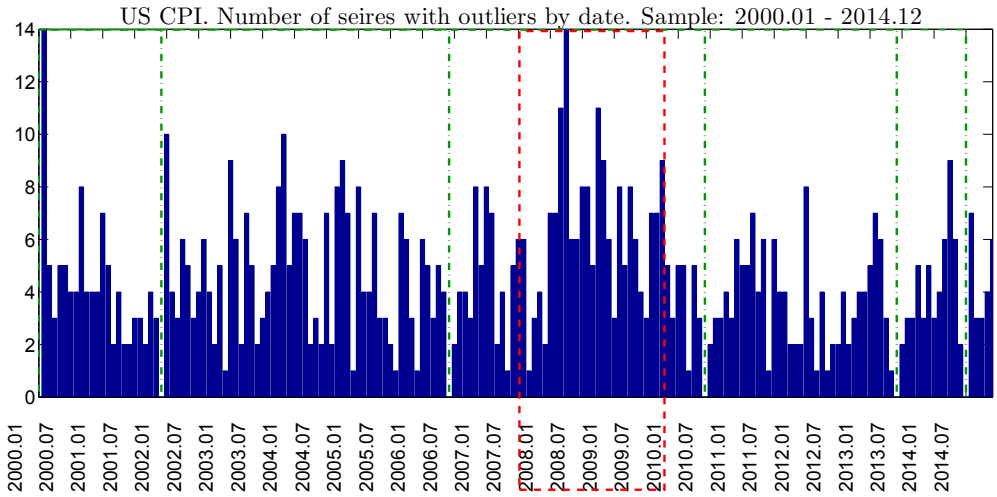
| | L+ | S+ | S- | L- | **Mean** | **Mean (% of T)** |
|---|---|---|---|---|---|---|
| NPF (25) | 1.3 | 2.0 | 1.4 | 0.4 | **5.2** | **2.9%** |
| ENE (6) | 2.3 | 3.2 | 1.3 | 1.3 | **8.2** | **4.5%** |
| PF (38) | 0.6 | 1.9 | 0.9 | 0.2 | **3.7** | **2.0%** |
| MAN_dur (51) | 0.5 | 1.6 | 1.3 | 0.6 | **3.9** | **2.2%** |
| MAN_NoDur (10) | 0.3 | 1.5 | 1.1 | 0.6 | **3.5** | **1.9%** |
| Serv (34) | 2.3 | 2.3 | 0.8 | 1.1 | **6.5** | **3.6%** |
| TOTAL(164) | 7.4 | 12.4 | 6.9 | 4.2 | **4.7** | **2.6%** |
| **PROP.** | **24%** | **40%** | **23%** | **13%** | **100%** | |

Numbers in parenthesis after the category name indicate the number of series in the category.
L+: Large (larger than $4\sigma$) and positive outliers.
S+: Small (smaller than or equal to $4\sigma$) and positive outliers.
L-: Large and negative outliers.
S-: Small and negative outliers.

Another point of interest regarding the analysis of outliers is its distribution by dates. Fig-

[6]Note that this grouping is not perfect for a basic component could include prices belonging to two broad categories

ure III.4 shows the number of series with an outlier at each of the 192 months of the sample. As it shows, the distribution if far from uniform, with some months having 14 (8.5%) series with outliers and some others with none. Interestingly, there seems to be a concentration around years 2008-2009, the sub-prime crisis period (red box of the figure).



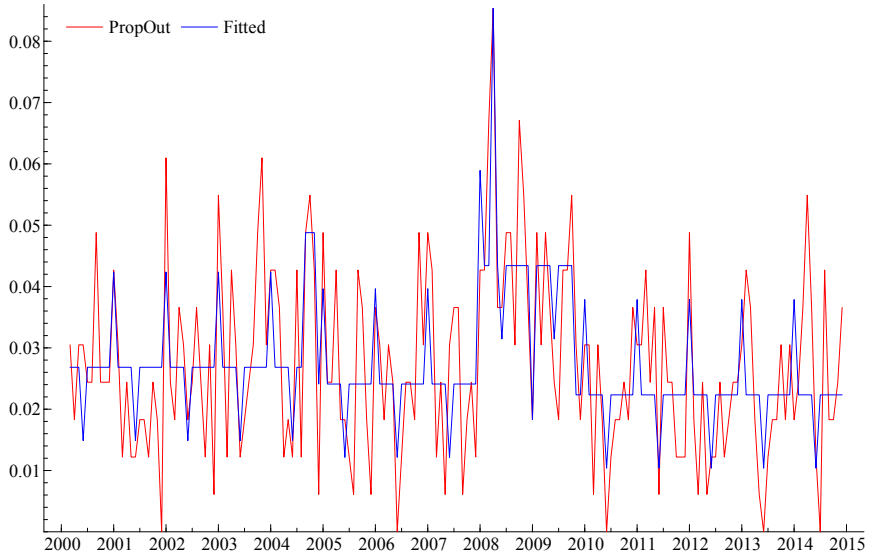

Note: Green doted lines indicate dates at which there are no contaminated series.

Figure III.4: Number of series with an outlier

To confirm that there is a mean shift in the number of contaminated components during the sub-prime crisis, we estimate a model for the proportion of series with outliers including as potential regressors lags 1 to 5, seasonal dummies and choosing the *Autometrics* option IIS+SIS[7] for impulses and steps detection with a target gauge equal to $1/T$. Results are summarized in figure III.5, from where five important conclusions can be drawn: i) There is seasonality in proportion of contaminated series; while in January there are, on average, more series with outliers, in June the proportion of series with outliers is reduced. ii) In 2004.11 there is a small reduction in the proportion of series with outliers with respect to previous years, that lasts until 2007.11. iii) In 2007.12 there is a significant and positive step that lasts until 2009.10. This confirms the observation about the sub-prime crises. iv) After 2009.10 the mean proportion of series with outliers is somewhat lower than before the crisis.

Finally it is noteworthy that the exhaustive outliers' search we made is not only relevant for individual series, but also for modeling the $CPI$ itself. This is so for outliers in the components are also outliers in the aggregate but, as we argue below, in many cases they can be estimated

[7]SIS stands for *Step Indicator Saturation*. The option IIS+SIS saturates the regression not only with impulses but also with steps, see Doornik et al. (2013)




| The estimation sample is: 2000(3) - 2014(12) | | | | | |
|---|---|---|---|---|---|
| | Coefficient | Std.Error | t-value | t-prob | Part.R^2 |
| Constant | 0.0226 | 0.0015 | 15.2000 | 0.0000 | 0.5773 |
| CSeasonal | 0.0155 | 0.0034 | 4.5700 | 0.0000 | 0.1101 |
| CSeasonal_5 | -0.0120 | 0.0032 | -3.7600 | 0.0002 | 0.0771 |
| I:2008(4) | 0.0420 | 0.0120 | 3.4900 | 0.0006 | 0.0672 |
| I:2009(1) | -0.0406 | 0.0124 | -3.2700 | 0.0013 | 0.0594 |
| S1:2004(8) | -0.0219 | 0.0070 | -3.1500 | 0.0019 | 0.0555 |
| S1:2004(11) | 0.0247 | 0.0071 | 3.5000 | 0.0006 | 0.0676 |
| S1:2007(12) | -0.0193 | 0.0033 | -5.9300 | 0.0000 | 0.1721 |
| S1:2009(10) | 0.0211 | 0.0030 | 6.9800 | 0.0000 | 0.2239 |

Notes:

- Steps ($S1$ :) take the value one from the first observation until the date indicated in the name of the step, and zero from then on.
- CSeasonal (Cseasonal_5) is the centered seasonal variable corresponding to January (June). It takes the value $1 - \frac{1}{12}$ in January (June) and $-\frac{1}{12}$ otherwise.

Figure III.5: Changes in the mean proportion of series with at least one outlier

only in the components. In order to use the individual outliers in a model for the $CPI$, we construct the *aggregated outlier* series ($AggOut$) as the weighted sum of all individual outliers multiplied by their coefficients, and include this series in a model for the $CPI$. Since the individual outliers will enter the $CPI$ weighted by the corresponding component's weight, we expect the coefficient of the *aggregated outlier* not to differ significantly from one.

The reason for expecting a unitary coefficient for the *aggregated outlier* can be easily seen by writing:

$$CPI_t = \sum_{i=1}^{N} w_{it} C_{i,t}, \tag{III.6}$$

where $C_{it}$ represents component $i$ at period $t$ and $w_{it}$ its weight. Now, the components can be expressed as the sum of their *core* plus their outliers:

$$C_{i,t} = C^*_{i,t} + \sum_{j=1}^{Q_i} \gamma_{ij} O_{ij,t} \equiv C^*_{i,t} + \sum_{j=1}^{Q_i} O^*_{ij,t}, \tag{III.7}$$

where $Q_i$ is the number of outliers in component $i$; $O_{ij,t}$ is the variable representing the j-*th* outlier of component $i$ and takes values 1 or 0; $\gamma_{ij}$ its coefficient; and $O^*_{ij,t} = \gamma_{ij} O_{ij,t}$.

The *aggregated outlier* series is defined as:

$$AggOut_t = \sum_{i=1}^{N} w_{it} \sum_{j=1}^{Q_i} O^*_{ij,t} \tag{III.8}$$

Plugging eq. (III.7) and in eq. (III.6) and using eq. (III.8) we get:

$$CPI_t = \sum_{i=1}^{N} w_{it} (C^*_{i,t} + \sum_{j=1}^{Q_i} O^*_{ij,t}) = \sum_{i=1}^{N} w_{it} C^*_{i,t} + AggOut_t \tag{III.9}$$

Equation (III.9) implies that $AggOut_t$ will have an unitary coefficient in a model for the $CPI$.

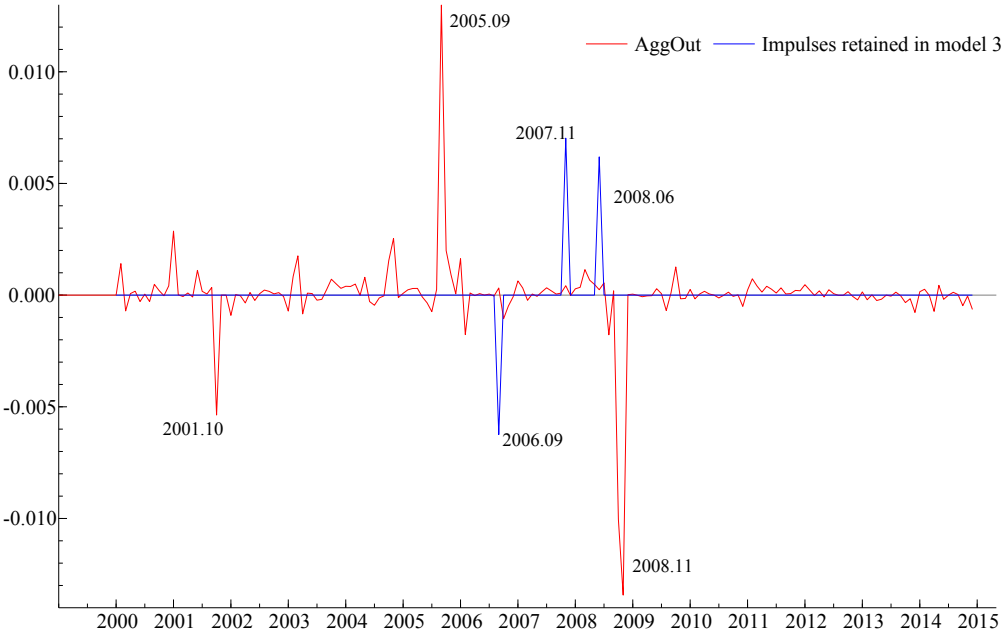
For assessing the usefulness of this variable to model the CPI, we compare three simple alternatives. Starting from the $GUM$; $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$, where $S_{it}$ is a centered seasonal dummy, we consider three possibilities to be estimated with *Autometrics*: (1) *Only IIS:* IIS is applied in previous GUM. (2) *Only AggOut:* The GUM is augmented with the series of $AggOut_t$ (IIS is not used). (3) *AggOut:* IIS is applied in the augmented GUM.

Table III.5 includes model selection criteria for the three possibilities. As it shows, the two models including $AggOut_t$ outperform model (1). Interestingly, model (3) seems to be the best option. This last result suggests two conclusions: some components' outliers — which are also outliers of the $CPI$ — are not identifiable in the model for the aggregate, and some CPI's outliers — which must be present in some component — are not identifiable in component's models, probably because these observations correspond to small outliers of the same sign in more than one component. Figure III.6 includes the *aggregated outlier* series and the impulses retained in model (3). Blue peaks represent the outliers that are not identifiable in disaggregated series (2006.09, 2007.11 and 2008.06).

Table III.5: Comparison of different models for the CPI

| | **AIC** | **SIC** | ***Adj.R²*** |
|---|---|---|---|
| **Only IIS** | -9.14 | -8.90 | 0.65 |
| **Only AggOutl** | -9.22 | -9.02 | 0.67 |
| **AggOutl + IIS** | -9.33 | -9.06 | 0.71 |

Basic *GUM*: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$.
*Only IIS:* IIS is applied in previous GUM.
*Only AggOutl:* The GUM is augmented with the series of $AggOut_t$ (IIS is not used).
*AggOutl:* IIS is applied in the augmented GUM.




- *GUM*: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + AggOut_t + \epsilon_t$.
- IIS is applied in previous GUM and the resulting impulses are added up using their coefficients (blue line).

Figure III.6: Aggregated and remaining outliers in $\Delta CPI$

To conclude the outlier analysis, table B.3 of appendix B.2 contains the estimated model (i). The *p-value* for the null that $AggOut's$ coefficient is equal to 1 is 0.27 thus, as expected, it is not rejected.

### III.6.3 Results of the pairwise tests

Since the pairwise approach does not deal with seasonal unit roots, we performed previous OSCB (see Osborn et al. (1988)) tests to all the components. Results indicate that they do not show seasonal unit roots in general and that the assumption of only one regular unit root and linear growth seems sensible (details are available upon request).

For the outlier corrected series (see §III.2), Johansen's tests are performed at the 5% of

significance and the number of lags for each pair is determined with the $AIC$, in a model without trend in the 'long run' and one common unit root restriction. Centered seasonal dummies are included in all models.

Among the 164 components there are 13366 possible pairs. Without considering the last three quality conditions of §III.4.2, we find 2333 pairs with a common unit root (the first condition — no deterministic trend — is always applied). After disregarding those that do not satisfy the other three conditions (significant coefficients, root not close to one and stability) we keep 853 pairs that are considered to satisfy stable and statistically sensible unit root restrictions.

For grouping the components by blocks we consider the strategy summarized in §III.5 augmented by the *relaxation procedure* also described in that section. Blocks with less than four series are disregarded to avoid spurious grouping (see chapter II). In the *relaxation step* a maximum of 2 *holes* is admitted[8] and the significance level augmented to 10%.

We also consider the three extensions mentioned in §III.4.1 (a weakly exogenous variable, parsimonious long lag and a special form of non-linearity). As exogenous variable we use the Real Effective Exchange Rate (constructed by the Bank for International Settlements[9], using consumer price indexes for trading partners to deflate nominal effective exchange rate). The inclusion of non-linearities was the only one that leaded to lose common unit root restrictions so we finally did not use it.

Table III.6 summarizes the results with and without the extensions of §III.4.1. As it shows, the extensions lead to discover 'long-run' relationships that were not found in the baseline case. Only by admitting Parsimonious Long Lag structures the proportion of series included in some block increases from 18 to almost 34. Regarding the $REER$, table III.6 suggests that including it when it is not needed may distort the results (compare the number of series in the second and third line of the table). This finding is in line with the simulations results in chapter II and those of Lütkepohl et al. (2003) and Johansen (1995) about that cointegration test's power decreases with the number of stochastic trends in the system. For this reason we prefer the sequential procedure in which $REER$ is included only in a second step after having constructed the blocks without considering this variable.

Note however that the direct strategy for including the $REER$ captures a significant larger weight of the CPI than the sequential procedure (compare the third column of table III.6 for

[8] If the original set has less than 6 series only one hole is admitted.
[9] https://www.bis.org/statistics/eer/index.htm

Table III.6: Summary Pairwise procedure results: total proportion of series and weight of all the (almost) fully cointegrated subsets with outlier corrected series

Total proportion of series and weight of all the *(almost) fully cointegrated subsets*

| | Num of Sets | Num of Sers | Weight(%)* | Weight top three(%)** |
|---|---|---|---|---|
| Without any extension | 4 | 18 | 4.0 | 1.3 |
| With PLL | 8 | 34 | 7.8 | 1.8 |
| With PLL and REER | 7 | 30 | 32.4 | 27.6 |
| With PLL and REER (sequential)[a] | 10 | 42 | 11.2 | 3.4 |

(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the CPI.
(**) Weight of the tree series with larger weights.
*(a)* In this strategy blocks are firstly formed without *REER* and then the procedure with *REER* is executed only for the series not included in any previous block.

the last two rows). As the last column of the table shows, this difference is explained by the inclusion of few 'heavy' components in the direct procedure. In fact, the inclusion of *'Owners' equivalent rent of primary residence'* (that represents 25% of the weight of the series we are dealing with) explains all the difference. We therefore still prefer the sequential procedure.

The last remark regards the outlier treatment. Results commented up to now were for the outlier corrected series. To assess the impact of this correction, table III.7 replicates table III.6 but without outlier correction. A noteworthy difference between the two tables is that when series are not corrected for outliers the number of series included in some block is systematically larger. This result may be due to two issues: co-breaking relationships may be 'confused' with common unit roots restriction in not corrected series, and power problems of the *GLS* procedure may lead to incorrectly find no common unit roots restrictions in too many pairs.

Since the outlier correction strategy does not distort the *gauge* of the pairwise method (see simulation results of table III.1) the 42 series of table III.6 may be considered as lower bound of the number of series that would be found without outlier treatment in case the components were not contaminated. Hence, the number of series in table III.7 can be *correctly* larger only if power issues of the GLS method were the main source of the differences between the two tables. In this case, one should expect that series in table III.7 include a large proportion of those in table III.6 plus some other series. However, when comparing the individual blocks obtained with both procedures coincidences are minor. For this reason and for on theoretical basis the appropriate approach is the one with outlier correction, we select the results of table III.1 for our application.

Table III.7: Summary Pairwise procedure results: total proportion of series and weight of all the (almost) fully cointegrated subsets with series *not* corrected for outliers

Total proportion of series and weight of all the *(almost) fully cointegrated subsets*

| | Num of Sets | Num of Sers | Weight(%)* | Weight top three(%)** |
|---|---|---|---|---|
| Without any extension | 6 | 34 | 22.1 | 12.2 |
| With PLL | 9 | 43 | 15.0 | 6.6 |
| With PLL and REER | 9 | 39 | 11.8 | 4.2 |
| With PLL and REER (sequential)[a] | 11 | 51 | 18.5 | 6.6 |

(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the CPI.
(**) Weight of the tree series with larger weights.
*(a)* In this strategy blocks are firstly formed without $REER$ and then the procedure with $REER$ is executed only for the series not included in any previous block.

#### III.6.3.1 Some detailed results

Table III.8 gives some details of the outcome for the procedure with $PLL$, *sequential* $REER$ and outliers correction. To make the table legible we use the same six broad categories as in table III.4.

The main conclusion of the table is that blocks of series sharing one common unit root (or two for block 10 and 11 which have $REER$) cannot be assigned to a single broad category. However, in almost all the cases, more than 85% of the *Set* weight is explained by two broad categories. The exceptions are blocks 6 and 7 for which the two most important categories explain 70% of the block's weights.

This observation has two relevant implications: first the ad-hoc method proposed by Boivin and Ng (2006) for extracting non-pervasive common factors, based on pre-grouping the series in broad categories, would not work for the US CPI; second, although a 'labeling' strategy that matches blocks with single broad categories is not possible, this could be done using just two categories.

### III.6.4 Forecasting the US CPI and all its components

See chapter VI.

Table III.8: Detailed results of the Pairwise procedure with common unit roots: number of series and proportion of weight by broad categories and blocks

| | MAN dur | | PF | | SERV | | NPF | | MAN No dur | | ENE | | Tot W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q | W | Q | W | Q | W | Q | W | Q | W | Q | W | |
| TOTAL | 51 | 17.2 | 38 | 12.6 | 34 | 51.0 | 25 | 4.8 | 10 | 3.8 | 6 | 10.7 | |
| Block 1 | 0 | 0.0 | 3 | 62.2 | 0 | 0.0 | 2 | 37.8 | 0 | 0.0 | 0 | 0.0 | 1.6 |
| Block 2 | 1 | 11.8 | 2 | 43.5 | 0 | 0.0 | 2 | 44.7 | 0 | 0.0 | 0 | 0.0 | 0.7 |
| Block 3 | 2 | 56.7 | 1 | 7.5 | 0 | 0.0 | 1 | 35.7 | 0 | 0.0 | 0 | 0.0 | 1.1 |
| Block 4 | 1 | 7.0 | 0 | 0.0 | 0 | 0.0 | 2 | 60.1 | 0 | 0.0 | 1 | 32.9 | 0.8 |
| Block 5 | 0 | 0.0 | 1 | 10.9 | 0 | 0.0 | 2 | 30.1 | 1 | 59.0 | 0 | 0.0 | 0.6 |
| Block 6 | 1 | 16.0 | 1 | 50.9 | 1 | 21.4 | 1 | 11.6 | 0 | 0.0 | 0 | 0.0 | 0.8 |
| Block 7 | 1 | 29.3 | 0 | 0.0 | 2 | 41.9 | 0 | 0.0 | 1 | 28.7 | 0 | 0.0 | 1.0 |
| Block 8 | 2 | 14.1 | 0 | 0.0 | 1 | 52.7 | 0 | 0.0 | 1 | 33.2 | 0 | 0.0 | 1.2 |
| Block 9 | 2 | 92.0 | 1 | 2.9 | 0 | 0.0 | 1 | 5.1 | 0 | 0.0 | 0 | 0.0 | 2.4 |
| Block 10 | 0 | 0.0 | 2 | 48.0 | 0 | 0.0 | 1 | 14.4 | 1 | 37.6 | 0 | 0.0 | 1.0 |
| TOTAL | 10 | | 11 | | 4 | | 12 | | 4 | | 1 | | 11.2 |

Columns $Q$ indicate the amount of series in each category and Set.
Columns $W$ indicate the total weight of each category in the CPI and the proportion of the weight of each category in each Set.
Last column contains the total weight of the blocks.

## III.7 Concluding Remarks

In this chapter we proposed a robustification strategy for the pairwise approach to discover common trends proposed in chapter II, and applied it in a real data exercise. The robustification includes a strategy for dealing with data irregularities and with short samples issues as well as some extensions to the design of the pairwise cointegration tests.

The outliers' treatment combines the Impulse Indicator Saturation (IIS) methodology (see Santos et al. (2008)) with the feasible GLS procedure proposed by Saikkonen and Lütkepohl (2000) to test cointegration in multivariate systems without the need of simulating new critical values. Outliers' dates are estimated by IIS and then these dates are used in the GLS procedure.

The outliers' correction strategy was analyzed in a simulation study. We found that using it when it is not required deteriorates the procedure's performance but not dramatically. Additionally, when outliers' dates are known, the pairwise approach behaves similarly to the case when outliers are not present (and no treatment is applied). When outliers' dates have to be estimated, relevant *potency* reductions are observed. In relation with this issue we highlighted the importance of correctly specifying the dynamic structure of the models in which the outliers' dates are estimated. Since, for simplicity, in the simulation exercise we did not selected the lag

structure in the IIS regressions, our results about *potency* must be considered as lower bounds for those that will be obtained when selecting the lags.

Comparing the results of correcting *vs.* not correcting for outliers in contaminated series we found that, although we have only lower bound potencies for the cases of estimated dates, estimating the dates is generally better than doing nothing.

Regarding small samples, we proposed a correction strategy and studied it by Monte Carlo. Results show that it provides significant *potency*[10] improvements, at the cost of a somewhat larger *gauge*[11]. This trade-off can be managed by what we called the *relaxation parameter* ($\lambda$), which determines the number of cointegration tests for which the nominal level is relaxed to a higher value. The larger $\lambda$, the larger the *potency* improvements and cost in terms of *gauge*. The user can select $\lambda$ according to her preferences. In the application we have used lambda equal two.

When applying the procedure to the US CPI we found that the groups generated by the pairwise procedure cannot be fully assigned to a single broad official category of prices but, in almost all the cases, more than 85% of the weight of the subset of fully cointegrated components is explained by two of them.

In regards to the analysis of outliers we found that the sub-prime crises is characterized by a significant and positive mean shift in the proportion of components with outliers. Additionally, we found that the aggregated outlier — a series constructed by aggregating the outliers of the components — helps to improve the model of the aggregate. This is because the are some outliers that are not identifiable in the aggregated series.

[10] The retention frequency of relevant variables.
[11] The retention frequency of irrelevant variables.

# Appendix B

# Appendix to chapter III

## B.1 The problem of consecutive outliers

Let $y_t$ be one component of the aggregate and assume that it is an $I(1)$ process with some level breaks such that;

$$\begin{aligned} y_t &= x_t + \gamma LS_t \\ \Phi(L)x_t &= \epsilon_t \end{aligned} \tag{B.1}$$

where $\Phi(L)$ is a polynomial in $L$ with one unit root; $\epsilon_t$ is a Gaussian white noise; $\gamma$ is $1 \times k$ vector of coefficients; $LS_t$ a $k \times 1$ vector of level shifts; and $x_t$ is the uncontaminated (unobserved) process. From (B.1):

$$\Phi(L)y_t = \Phi(L)\gamma LS_t + \epsilon_t \tag{B.2}$$

If $\Phi(L) = (1 - L)$, then (B.2) becomes;

$$\Delta y_t = \gamma \Delta LS_t + \epsilon_t$$

In this simple case, the IIS methodology will, hopefully, find the correct break dates and the strategy outlined in section III.2.2 will, hopefully, work. However if, as is usually the case, $\Phi(L)$ has a more complex structure, the strategy previously outlined needs more elaboration. Let, for instance, $\Phi(L) = (1 - L)(1 - \phi_1 L)$. Then, the 'true' model for $\Delta y_t$ becomes:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \gamma \Delta LS_t - \phi_1 \gamma \Delta LS_{t-1} + \epsilon_t, \tag{B.3}$$

and in the more general case that $\Phi(L)$ has a larger structure:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + ... + \phi_p \Delta y_{t-p} + \gamma \Delta LS_t - \phi_1 \gamma \Delta LS_{t-1} - \phi_2 \gamma \Delta LS_{t-2} - ... - \phi_p \gamma \Delta LS_{t-p} + \epsilon_t.$$

Therefore, applying IIS we may find more than $k$ impulses, so that the break dates do not emerge directly, for we cannot distinguish between lagged values of a break and different breaks. This is a relevant issue since the GLS procedure requires the inclusion of $LS_t$ only.

One possible solution for this issue may be to apply IIS in a model with AR residuals (instead of the lagged dependent variable). However, such a model includes non-linear restrictions in the parameters, what would make the estimation procedure highly complex given the large amount of indicators we need to include.

In order to keep the estimation procedure simple we proceed as follows. First, apply IIS with the lagged dependent variable as regressors and store the retained impulses. Next, if there are potential lagged impulses, estimate a model with AR residuals including the retained impulses as regressors, test the significance of the potential lagged impulses' parameters and drop the insignificant ones. With this simple procedure we are testing if the coefficients of the lagged impulses satisfy the restrictions implied by the dynamics of the model, or are *true* impulses.

To see why this procedure should work, assume that we are dealing with an AR(1) model so that the true model is given by expression (B.3). Assume also that after IIS we retained $LS_t$ and $LS_{t-1}$, thus, the procedure requires to estimate:

$$\Delta y_t = \lambda_1 \Delta LS_t + \lambda_2 \Delta LS_{t-1} + w_t,$$
$$w_t = \phi_1 w_{t-1} + \epsilon_t,$$

from where;

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \lambda_1 \Delta LS_t + (\lambda_2 - \phi_1 \lambda_1) \Delta LS_{t-1} - \phi_1 \lambda_2 \Delta LS_{t-2} + \epsilon_t.$$

Then, we need to test $(\lambda_2 - \phi_1 \lambda_1) = 0$. If the true model is (B.3) we will not reject the hypothesis and keep only $\Delta LS_t$ as the *true* impulse. Note that this argument also works for a general $AR(p)$ model and any distribution of the impulses.

## B.2 Performance of *1-cut IIS* for series in §III.2.3

Table B.1: Average Gauge and Potency (%) of *1-cut* IIS applied to the 100 series in each scenario

| | **Scenario 1**, ($n_1 = 10$) | | **Scenario 3**, ($n_1 = 25$) | |
|---|---|---|---|---|
| | Gauge | Pot | Gauge | Pot |
| Number of breaks: 1 | | | | |
| $\gamma$ =4 | 0.38 | 83.1 | 0.39 | 79.1 |
| $\gamma$ =5 | 0.38 | 96.6 | 0.38 | 94.4 |
| Number of breaks: 2 | | | | |
| $\gamma$ =4 | 0.39 | 78.6 | 0.40 | 74.7 |
| $\gamma$ =5 | 0.40 | 93.9 | 0.40 | 91.6 |
| Number of breaks: 3 | | | | |
| $\gamma$ =4 | 0.40 | 73.6 | 0.39 | 68.8 |
| $\gamma$ =5 | 0.32 | 90.4 | 0.32 | 86.6 |

- This table describes the performance of *1-cut* IIS when applied to the series used in section III.2.3.2.
- Figures are averages across series and the 1000 experiments.
- See §III.2.3 for the details of the Monte Carlo design.
- Gauge measures the retention frequency of irrelevant impulses.
- Potency measures the retention frequency of relevant impulses.

## B.3 Inclusion of a weakly exogenous variable in the sub-systems in which cointegration is tested

Notation reference: processes labeled as $\varpi_t$ are stationary, but not necessarily white noise.

### B.3.1 General Framework

Let $Y_t = (X'_{1,t}, X'_{2,t}, EXO'_t)'$, where $X_i$ represents component $i$. The three dimensional models are:

$$(I_n - \Pi_1 L - ... - \Pi_k L^k) Y_t = \mu_t + \epsilon_t, \tag{B.4}$$

where, $e_t$ is a $M \times 1$ vector of Gaussian white noises.

If the system is cointegrated:

$$\Delta Y_t = \mu_t + \alpha\beta' X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1}\Delta X_{t-k-1} + \epsilon_t, \tag{B.5}$$

where $\alpha\beta' \equiv \Pi = \sum_{i=1}^{k} \Pi_i - I$, and $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$

From the Granger Representation theorem:

$$Y_t = X_0 + C(1)\sum_{i=1}^{t} \epsilon_t + C(1)\mu t + C^*(L)\epsilon_t, \tag{B.6}$$

where $C(1) = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp$, with $\Psi = I_n - \sum_{i=1}^{k-1} \Phi_i$. The common stochastic trends in the system are, $\alpha'_\perp \sum_{i=1}^{t} \epsilon_t$.

Since we will focus on cases where $EXO$ is weakly exogenous (last row of $\alpha$ is a zero vector), the vector $[0, 0, 1]$ is orthogonal to $\alpha$, so that $\sum_{i=1}^{t} \epsilon_{exo,t}$ is a common stochastic trend. If the system has two common trends, the other one would be a linear combination of $\sum_{i=1}^{t} \epsilon_{1,t}$ and $\sum_{i=1}^{t} \epsilon_{2,t}$.

When including $EXO$ in the sub-systems there are four possible outcomes of the cointegration tests:

a. There are no cointegration relations ($r = 0$).

b. There is only one cointegration relationship ($r = 1$).

c. There are only two cointegration relationships ($r = 2$).

d. There are three cointegration relationships (all the variables are stationary, $r = 3$).

Since cases (a) and (d) are not considered in the construction of the 'fully cointegrated' subsets, we focus in cases (b) and (c).

Restricting our attention to cases (b) and (c), the general situation is that there will be some systems with $r = 1$ and some others with $r = 2$. In the first case, there will be two common trends; one determined by the cumulated shocks of EXO, and the other by a combination of the cumulated shocks of the two prices. When $r = 2$, there will be only one common trend generated by the cumulated shocks of EXO.

### B.3.2 'Mixed' sub-systems

We now show that, for one series, finding $r = 1$ and $r = 2$ in two different sub-systems is incompatible.

Consider the case that for the sub-system $(X_1, X_2, EXO)$ we find $r = 2$ (call this as *sub-system c*) but for $(X_1, X_3, EXO)$ we find $r = 1$ (*sub-system b*).

From the *sub-system c* we can write:

$$X_{1t} = \theta^c_{11} CT_{exo,t} + \varpi^c_{1t}$$
$$X_{2t} = \theta^c_{21} CT_{exo,t} + \varpi^c_{2t}.$$

Similarly, from *sub-system b*:

$$X_{1t} = \theta^b_{11} CT_{x,t} + \theta^b_{12} CT_{exo,t} + \varpi^b_{1t}$$
$$X_{3t} = \theta^b_{21} CT_{x,t} + \theta^b_{22} CT_{exo,t} + \varpi^b_{3t}$$

Equalizing the first equations of the two sub-systems:

$$\theta^c_{11} CT_{exo,t} + \varpi^c_{1t} = \theta^b_{11} CT_{x,t} + \theta^b_{12} CT_{exo,t} + \varpi^b_{1t},$$

from where:

$$(\theta^c_{11} - \theta^b_{12}) CT_{exo,t} = \theta^b_{11} CT_{x,t} + (\varpi^b_{1t} - \varpi^c_{1t}). \tag{B.7}$$

Without any restriction in the coefficients, expression (B.7) is a contradiction because it implies cointegration between two independent random walks.

For equation (B.7) to be *possible* two conditions are required:

(i) $\theta^c_{11} = \theta^b_{12}$, *and*

(ii) $\theta^b_{11} = 0$

The two conditions are required at the same time because with only one of them a unit root process would be equal to a stationary one. Note that condition (ii) requires $\beta_{11} = 0$ in *system b*. Thus, if we require the cointegration coefficients associated to the prices to be different from zero, sub-systems $b$ and $c$ cannot happen at the same time.

This result implies that when $EXO$ is included (and is weakly exogenous) the search of 'fully cointegrated' subsets should not mix sub-systems with different number of cointegration relationships.

When sub-systems have two cointegration relationships (one common trend) we are in the 'regular' case and 'fully cointegrated' sets can be formed as usual.

We now consider the case where the sub-systems have only one cointegration relationship.

### B.3.3 Sub-systems with r=1

Consider the following two sub-systems:

$$\begin{aligned} X_{1t} &= \theta^1_{11} CT_{x_{12},t} + \theta^1_{12} CT_{exo,t} + \varpi^1_{1t} \\ X_{2t} &= \theta^1_{21} CT_{x_{12},t} + \theta^1_{22} CT_{exo,t} + \varpi^1_{2t} \end{aligned} \tag{B.8}$$

$$\begin{aligned} X_{1t} &= \theta^2_{11} CT_{x_{13},t} + \theta^2_{12} CT_{exo,t} + \varpi^2_{1t} \\ X_{3t} &= \theta^2_{21} CT_{x_{13},t} + \theta^2_{22} CT_{exo,t} + \varpi^2_{3t} \end{aligned} \tag{B.9}$$

From the first line in (B.9):

$$CT_{x_{13},t} = \frac{X_{1t} - \theta^2_{12} CT_{exo,t} - \varpi^2_{1t}}{\theta^2_{11}} \tag{B.10}$$

Plugging (B.10) in the second line of (B.9):

$$\begin{aligned} X_{3t} = \frac{\theta^2_{21}}{\theta^2_{11}}(X_{1t} - \theta^2_{12} CT_{exo,t} - \varpi^2_{1t}) + \theta^2_{22} CT_{exo,t} + \varpi^2_{3t} \\ X_{3t} = \lambda X_{1t} - \kappa CT_{exo,t} + e_{3t}, \end{aligned} \tag{B.11}$$

where $\lambda = \frac{\theta^2_{21}}{\theta^2_{11}}$; $\kappa = \theta^2_{22} - \frac{\theta^2_{21}\theta^2_{12}}{\theta^2_{11}}$; and $e_{3t} = \varpi^2_{3t} - \frac{\theta^2_{21}}{\theta^2_{11}}\varpi^2_{1t}$.

Now, using the first equation of sub-system (B.8), we get:

$$\begin{aligned} X_{3t} = \lambda(\theta^1_{11} CT_{x_{12},t} + \theta^1_{12} CT_{exo,t} + \varpi^1_{1t}) - \kappa CT_{exo,t} + e_{3t} \\ X_{3t} = \gamma CT_{x_{12},t} + \delta CT_{exo,t} + e^*_t, \end{aligned} \tag{B.12}$$

where $\gamma = \frac{\theta^2_{21}\theta^1_{11}}{\theta^2_{11}}$; $\delta = \frac{\theta^2_{11}\theta^1_{12} - \theta^2_{22}\theta^2_{11} + \theta^2_{21}\theta^2_{12}}{\theta^2_{11}}$; and $e^*_t = \varpi^2_{3t} + \frac{\theta^2_{21}}{\theta^2_{11}}(\varpi^i_{1t} - \varpi^2_{1t})$

Therefore, provided that $\theta^2_{11} \neq 0$, $X_{3t}$ can also be written as a linear combination of the two common trends in sub-system (B.8) plus a stationary component. Thus, 'fully cointegrated'

subsets can be formed considering sub-systems with only one cointegration relationship. The only difference with the original case is that in these sets there are two common trends, one determined exclusively by $EXO$ and the other by the components.

The condition $\theta_{11}^2 \neq 0$ means that the stochastic trend $CT_{x_{13},t}$ is present in $X_{1,t}$, which implies that $X_{1t}$ is present in the cointegration relation. Thus, sub-systems in which the cointegration relationship does not include both components are to be disregarded.

### B.3.4 Empirical Strategy

1. Focus first on sub-systems with $r = 2$ and look for 'fully cointegrated' subsets.

2. Disregard series in some 'fully cointegrated' subsets and repeat previous step but now considering sub-systems with $r = 1$.

Note that the order of the two steps should not matter given the incompatibility of 'mixed' sub-systems.

#### B.3.4.1 The forecasting equations

**Series in sets with r=2**

When all the sub-systems have $r = 2$ we have only one common trend and $n_1$ cointegration relations (for we have $n_1$ components plus $EXO$).

Any cointegrated system can always be normalized such that: $[\beta' = \tilde{\boldsymbol{\beta}}'_{r\times(N-r)}, I_r]$. Therefore, the $n_1$ cointegration relationships can be expressed as $n_1 - 1$ deviations of the components with respect to one of them, and the deviations of $EXO$ with respect to that same component. These $n_1$ cointegration relationships are to be included as potential regressors in the forecasting equations.

**Series in sets with r=1**

As showed above, a subset in which all 'augmented pairs' have $r = 1$ is a system with two common trends. Since we have $n_1$ components plus $EXO$, there are $n_1 + 1 - 2 = n_1 - 1$ cointegration relations. Using the normalization mentioned above, the full system can be written as follows:

$$\begin{bmatrix} \Delta X_{1,t} \\ \Delta X_{2,t} \\ \vdots \\ \Delta X_{n_1,t} \\ \Delta EXO_t \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1,n_1-1} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2,n_1-1} \\ \alpha_{n_1,1} & \alpha_{n_1,2} & \dots & \alpha_{n_1,n_1-1} \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & 1 & 0 & \dots & 0 \\ \beta_{21} & \beta_{22} & 0 & 1 & \dots & 0 \\ \vdots & & & & & \\ \beta_{n_1-1,1} & \beta_{n_1-1,2} & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{n_1,t-1} \\ EXO_{t-1} \end{bmatrix} + \begin{bmatrix} \varpi_{1,t} \\ \varpi_{2,t} \\ \vdots \\ \varpi_{n_1,t} \\ \varpi_{EXO,t} \end{bmatrix} \tag{B.13}$$

This formulation means that cointegration relationships can be written as deviations of all — but two — components with respect to two of them, and the deviation of the $EXO$ with respect to those two components. These $n_1 - 1$ cointegration relationships are to be included as potential regressors in the forecasting equations.

A possibility could be to select $EXO$ as one of the 'normalizing' series, but since we are not testing the significance of $EXO$ in the cointegration relationships, this strategy could be risky.

## B.4 Additional Results for the analysis of outliers

Table B.2: Series with 5% or more outliers

| 1 | | **Weight** | **Cat** | **Tot. Num Outl** | **%** |
|---|---|---|---|---|---|
| 2 | Lettuce | 0.072 | NPF | 10 | 5.2% |
| 3 | Bananas | 0.087 | NPF | 12 | 6.3% |
| 4 | Propane, kerosene, and firewood | 0.097 | ENE | 13 | 6.8% |
| 5 | Fuel oil | 0.139 | ENE | 13 | 6.8% |
| 6 | Food from vending machines... | 0.064 | PF | 15 | 7.8% |
| 7 | Cable and satellite tv and radio... | 1.468 | MAN_dur | 10 | 5.2% |
| 8 | Other furniture | 0.128 | MAN_dur | 11 | 5.7% |
| 9 | Other appliances | 0.12 | MAN_dur | 11 | 5.7% |
| 10 | Telephone hardware, calculators, and... | 0.068 | MAN_dur | 12 | 6.3% |
| 11 | Photographic equipment and supplies | 0.058 | MAN_dur | 12 | 6.3% |
| 12 | Moving, storage, freight expense | 0.116 | Serv | 10 | 5.2% |
| 13 | Telephone services | 2.462 | Serv | 11 | 5.7% |
| 14 | Domestic services | 0.279 | Serv | 11 | 5.7% |
| 15 | Parking and other fees | 0.235 | Serv | 11 | 5.7% |
| 16 | Fees for lessons or instructions | 0.211 | Serv | 11 | 5.7% |
| 17 | Technical and business school tuition... | 0.039 | Serv | 11 | 5.7% |
| 18 | Financial services | 0.228 | Serv | 14 | 7.3% |
| 19 | Physicians' services | 1.59 | Serv | 16 | 8.3% |
| 20 | Tenants' and household insurance | 0.375 | Serv | 19 | 9.9% |
| 21 | State motor vehicle registration... | 0.312 | Serv | 32 | 16.7% |
| | **Total weight** | **8.15** | | | |

Table B.3: Estimated model *Only AggOut*

| | Coefficient | Std.Error | t-value | t-prob | $Part.R^2$ |
|---|---|---|---|---|---|
| $DLCPI_1$ | 0.3750 | 0.0559 | 6.7100 | 0.0000 | 0.2105 |
| $DLCPI_2$ | -0.2933 | 0.0516 | -5.6800 | 0.0000 | 0.1605 |
| $DLCPI_{12}$ | -0.1079 | 0.0554 | -1.9500 | 0.0530 | 0.0220 |
| **AggOut** | **1.1388** | **0.1265** | **9.0100** | **0.0000** | **0.3243** |
| $CSeason_2$ | 0.0030 | 0.0007 | 4.4600 | 0.0000 | 0.1055 |
| $CSeason_6$ | -0.0023 | 0.0007 | -3.4800 | 0.0006 | 0.0669 |
| $CSeason_8$ | -0.0017 | 0.0007 | -2.6200 | 0.0095 | 0.0391 |
| $CSeason_9$ | -0.0037 | 0.0007 | -5.2800 | 0.0000 | 0.1416 |
| $CSeason_{10}$ | -0.0042 | 0.0007 | -5.7100 | 0.0000 | 0.1616 |
| $CSeason_{11}$ | -0.0049 | 0.0007 | -6.4900 | 0.0000 | 0.1997 |
| Constant | 0.0019 | 0.0002 | 8.2700 | 0.0000 | 0.2881 |
| | | | | | |
| $\sigma$ | 0.00234232 | | | | |
| $R^2$ | 0.685407 | | | | |
| $AdjR^2$ | 0.666792 | | | | |
| AIC | -9.21619 | | | | |
| SIC | -9.02106 | | | | |

- Basic $GUM$: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$.
- IIS is not applied in previous GUM.

$CSeason$ is a centered seasonal dummy that takes the value $1 - 1/12$ in January and $-1/12$ otherwise. $CSeason_i$ is the $i^{th}$ lag of $CSeason$.

*This page intentionally left blank.*

# Chapter IV

# Discovering common cycles in a large set of disaggregates: statistical procedures and their properties

## IV.1 Introduction

The presence of commonalities in the short-run dynamics of macroeconomic variables has been extensively documented. Referring to the movements of macroeconomic variables around their trends, Lucas (1977) points out: "Output movements across broadly defined sectors move together,... Prices generally are pro-cyclical". Long Jr and Plosser (1987) develop a theoretical model to explain commovments in sectoral output, and Engle and Issler (1995) derive the reduced form of Long Jr and Plosser's (1987) model, and state the conditions for the existence of common cycles. After Engle and Kozicki's (1993) seminal article on testing common features, several empirical studies testing for short-run commonalities among macroeconomic series appeared in the literature. Engle and Kozicki (1993) themselves find international commonalities in GNP data of OECD counties; Vahid and Engle (1993) find common cycles among regional per capita incomes in the US; and Engle and Issler (1995) find common cycles for sectoral outputs of the US economy also. Using techniques for detecting common cyclical features, Candelon et al. (2005) study financial contagion during the 1997 Hong Kong stock market crises and find evidence of contagion. Hecq et al. (2006) find common cyclical features among the GDP of five Latin American countries; and Cubadda (2007) finds commonalities in the short run movements

of the monthly indicators that The Conference Board uses to build the composite coincident indicator of the business cycle in the US.

The study of common cyclical features in macroeconomic series is relevant not only for understanding the interrelations among them, but also for constructing better empirical models and obtaining more accurate forecasts. Vahid and Issler (2002) analyze the importance of the restrictions implied by common-cyclical features for forecasts, impulse-response functions, and variance-decomposition analysis of economic time series. As they argue, the reduction in the number of parameters of typical macroeconomic VAR models derived from the existence of common cycles can be substantial, and much larger than those implied by cointegration. Therefore, remarkable efficiency gains can be obtained by imposing — correct — common cycles restrictions, from which forecasting accuracy improvements would follow. In a Monte Carlo study Vahid and Issler (2002) confirm that reduced rank models can lead to significant forecasting accuracy improvements with respect to unrestricted models.

In addition to the large number of empirical applications, Engle and Kozicki (1993) motivated a number of extensions to their original definition of *serial correlation common feature* (SCCF). These extensions were devoted to solve some limitations of the original one, and to make it applicable in more general data settings. In this chapter, we review the different definitions of common cyclical features and their estimation and testing procedures. This literature review is constructed in a way that the advances in the definitions and procedures are connected in a natural logical way along the subsections of §IV.2.

Although the equivalence between *common cyclical features* and *common cycles* is valid only for the original SCCF, we will refer to common cycles in a broad sense that includes all types of common cyclical features.

The main objective of this chapter is to extend the pairwise procedure studied in chapters II and III for the consideration of common-cyclical features. That is, we design a strategy to discover subsets of components that share single common cyclical features, and to use the restrictions derived from those subsets to construct single-equation models for all the disaggregates. The strategy for discovering those subsets is similar as that of chapter II; we perform common cycles tests between all the $N(N-1)/2$ pairs that exist in a set of $N$ variables, and look for subsets in which all the series share the cycle.

This strategy is motivated by Espasa and Mayo-Burgos (2013), who apply a approach strategy

to find the largest subset of series that share a single SCCF, as defined by Engle and Kozicki (1993). However, our approach is more rigorous than Espasa and Mayo-Burgos's (2013) in several aspects: (1) We analytically justify the validity of proceeding in a pairwise fashion for discovering 'single-cycle' subsets. (2) We use a more general definition of common cycles and admit that the 'single-cycle' subsets can be overlapped with the 'fully cointegrated' ones (see chapter II). (3) By means of a number of Monte Carlo exercises, we study the properties of the pairwise approach applied to common cycles and compare its behavior with a DFM alternative in some specific frameworks. (4) Finally, we do not restrict the search to the largest subset but look for all possible subsets of series sharing a single cycle.

Regarding (1), we study the transitiveness of different common cyclical features definitions in the framework of the pairwise strategy, and show that the original SCCF is the only transitive one. This result will, in principle, invalidate the pairwise approach for other types of common cyclical features. We show, however, that after a simple generalization of the testing strategy, more general common cyclical features also become transitive. This result led us to the second point, as we can consider different commonalities from the considered by Espasa and Mayo-Burgos (2013), and allow for overlapped 'single-cycle' and 'fully cointegrated' subsets. Nonetheless, the generalization of the testing procedures bear some difficulties that required a careful study (see §IV.5.2).

Point (3) concerns the study of the pairwise procedure's properties by means of Monte Carlo exercises. Focusing on the problem of multiple testing, also studied in chapter II for the case of common trends, we argue that the inclusion of wrong series in the estimated 'single-cycle' subset is not an issue, asymptotically. Therefore, the main issue is the inclusion of a large proportion of the correct series. Unfortunately, we do not have a result similar to theorem 1 of chapter II to show that tests are asymptotically equivalent and rule out the multiple testing problem inside 'single-cycle' subsets. On the contrary, we show that tests inside those subsets may, under some conditions, be independent.

To study the degree of dependency among tests inside 'single-cycle' subsets, we define two possible sources of dependency: that derived from the fact that all the series inside a subset have the same cycle (*dependence a*); and that derived from the fact that we are testing all the pairs of a set of series, so that each series appears in more than one pair. That is, the test statistic for the pair $X_{1,t}, X_{2,t}$ may not be independent from that of $X_{1,t}, X_{3,t}$ (*dependence b*).

We show that, under some conditions, both types of dependence may be very low, what makes the probability of discovering the true 'single-cycle' subset a fast decreasing function of the subset's size. This is an undesirable property of our procedure.

We show, however, that a relaxation procedure — similar to that defined in chapter III — virtually solves this problem with almost zero cost, asymptotically. That is, after the relaxation procedure we will discover a large proportion of the true series in the single-cycle subset with high probability, regardless the size of the true subset, and with almost no cost in terms of including wrong series.

As we did in chapter II, we focus in the case of fixed $N$ and $T$ going to infinity. Nonetheless, in §IV.6 we give some preliminary arguments supporting that our theory can be extended for the case of both $T$ and $N$ going to infinity. We obtain a similar insight as for the case of common trends; we would not need the common cycles to be pervasive. That is, the size of the single-cycle subsets could grow at a lower rate than $N$. Additionally, we would be able to deal with cases where $N/T \to \infty$.

Another issue of interest for the pairwise strategy is its relative performance with respect to a full model strategy, when the number of series is small and the latter approach feasible. We carry out this comparison by Mote Carlo and conclude that, when common cycles are pairwise detectable, nothing is lost for proceeding by pairs with respect to the full model approach. On the contrary, significant power gains for discovering the true number of common cycles may emerge from the pairwise procedure.

For the comparison with DFM we carry out other Monte Carlo experiments in which we consider both, stationary and non-stationary DGPs. We analyze the results in the light of the theoretical connection between restricted VAR models and DFM. As general conclusion, Monte Carlo experiments show that the pairwise approach outperforms DFM in some situations of empirical interest.

The rest of the chapter is organized as follows. In §IV.2, we review the alternative definitions of common-cyclical features and their estimation and testing procedures. In §IV.3, we study the relationship between VAR models with common features and DFM. In §IV.4, we focus on the analysis of the properties of the pairwise approach, devoting §IV.4.1 to the study of transitivity, §IV.4.2 to the generalization for making common cycles definitions to be transitive, §IV.4.3 to the asymptotic properties, and §IV.4.4 for the comparison of the pairwise approach with

the full model strategy. In §IV.5, we study the problems raised by the generalization of the testing strategy and we carry out the comparison with DFM. Finally, in §IV.6 we give an initial discussion for the case of $N \to \infty$, and in §IV.7 we present the main conclusions.

## IV.2 Common stationary features, methodological issues

The literature on common cycles is somewhat more recent than that on cointegration but is based on a similar motivation. As cointegration implies the existence of common stochastic trends in non-stationary time series, the concept of common cycles is related to the existence of common short-run (or medium-run) co-movements between the stationary components of a multivariate time series. The cornerstone in this literature was put by Engle and Kozicki's (1993) seminal article. Since then, some relevant improvements in the definition and the estimation and testing procedures have been developed. In this section we give a literature review of the most relevant types of short-run comovements, including the procedures available to detect them.

A chronological story that tries to connect all the proposals in the common cycles literature is the basic thread that runs through the five subsections of this section. Except for the first one, each subsection is motivated by the limitations on the definitions and concepts introduced in the previous one. Therefore, apart from the literature review, the main value of this section is the connection of the main contributions in this literature.

### IV.2.1 Serial correlation common feature; the initial idea

Engle and Kozicki (1993) generalize the concept of common trends to other possible common features (serial correlation, heteroscdasticity, excess kurtosis), and state that a certain feature is said to be common if a non-zero linear combination of a multivariate time series fails to have the feature even though each of the series individually have it. They develop a general testing procedure for any class of common feature, and find a specific test statistic with known distribution for the *serial correlation common feature (SCCF)*, which is the feature of their major interest. Specifically, Engle and Kozicki (1993) define the presence of a *SCCF*, when a linear combination of serially correlated time series is an innovation with respect to the past of the series.

Assume that the data can be represented by a VAR in differences: $\Delta X_t = \sum_{i=1}^{k} \Pi_i \Delta X_{t-i} + \epsilon_t$. The existence of a *SCCF* requires the existence of a $(n \times s)$ full column rank matrix $\delta$ such that

$\delta'\Delta X_t$ does not present serial dependence on the past of $\Delta X_t$, what implies $\delta'\Pi_i = 0$, for all $1 \leq i \leq k$. In this case, $\delta$ contains the *serial common correlation vectors*. All the autoregressive matrices would have a left null space that includes $\delta$, so the rank of $\delta$ will be the rank of the left null space of $\Pi_i$. Thus, $\Pi_i$ will have rank $n - s$.

This reduced rank condition on the matrices $\Pi_i$ suggests the application of reduced rank regression methods as the applied by Johansen (1988) for cointegration. The same procedures are valid in this case, with the difference that we are now looking for zero canonical correlations, instead of maximum ones as in cointegration analysis. We are now looking for zero canonical correlations between $\Delta X_t$ and $[\Delta X_{t-1}, ..., \Delta X_{t-k}]$. The *SCCF* vectors will be the eigenvectors related to $\Delta X_t$ and associated with the zero eigenvalues, i.e., zero squared canonical correlations.

A test for the existence of $s$ zero canonical correlations ($n - s$ common serial correlation features) is given by Anderson (2003), and it can be used as a test for the dimension the cofeature space. Here we include the slight adaptation of the test provided by Vahid and Engle (1993):

$$C(p, s) = -(T - k - 1) \sum_{i=1}^{s} ln(1 - \lambda_i), \tag{IV.1}$$

where $\lambda_i$ $(i = 1, ..s)$ are the smallest $s$ eigenvalues in the canonical correlation problem; $T$ is the sample size; and $k$ is the number of lags in the model. Under the null (the smallest $s$ eigenvalues are zero), the statistic has a $\chi^2$ distribution with $(s^2 + snk + sr - sn)$ degrees of freedom. The reason for the degrees of freedom can be easily seen by thinking of this problem as a pseudo-structural system of equations and counting the number of restrictions (see Vahid and Engle (1993) for further details).

### IV.2.2 *SCCF* in I(1) cointegrated variables

As previously described, Engle and Kozicki (1993) develop their methodology for stationary variables, so that, in most of the cases, the analysis must be carried out for the differenced variables, leading to informational losses if cointegration relationships exist. Vahid and Engle (1993) extend the framework to I(1) cointegrated systems by proposing a procedure for estimating *SCCF* vectors given the existence of common trends. Interestingly, the authors show that the presence of *SCCF* among the first differences of cointegrated I(1) variables is equivalent to the existence of common cycles in the sense of Beveridge and Nelson (1981). This can be seen using the moving average representation of a cointegrated VAR model derived form the Granger

Representation Theorem, which ignoring deterministic terms and assuming zero initial values can be written as:

$$X_t = C(1)\sum_{i=1}^{t}\epsilon_i + C^*(L)\epsilon_t, \tag{IV.2}$$

where $\Delta X_t = C(L)\epsilon_t$; $C(L) = C(1) + \Delta C^*(L)$; and $C_i^* = \sum_{j>i} -C_j,\ \forall i$.

This representation splits the series into a non-stationary stochastic trend ($C(1)\sum_{i=1}^{t}\epsilon_i$) and a stationary component ($C^*(L)\epsilon_t$), called 'cycle'. The rank of $C(1)$ determines the existence of common trends. If $C(1)$ has not full rank, there exists a matrix $\beta$ such that $\beta' X_t$ is stationary, so that all trend components are removed from eq. (IV.2) because $\beta' C(1) = 0$. Therefore, as noted by Vahid and Engle (1993), the cointegration relations are linear combinations of the cycles.

Similarly, if there exists some linear combinations of $X_t$ which do not contain cycles, i.e., $\delta' C^*(L) = 0$, $\delta$ will be a 'common cycle' vector, and $\delta' X_t$ will be a linear combination of the trends. Now, if such a vector exists, it must be true that, $\delta' C_i^* = 0$ for all $i \geq 0$. But the coefficient matrices $C_i^*$ are:

$$\begin{aligned} C_i^* &= \sum_{j>i} -C_j,\ \forall i \quad \textit{which implies}: \\ C_{i+1} &= C_{i+1}^* - C_i^* \qquad \forall\, i \geq 0, \end{aligned} \tag{IV.3}$$

Hence, the condition $\delta' C_i^* = 0$, also implies that $\delta' C_i = 0$. This means that $\delta' X_t$ will be an innovation with respect to the past of $X_t$, so that $\delta$ is a *SCCF* vector (and the converse also holds). Therefore, the levels of the variables present common cycles if and only if its differences present common serial correlation. This is the reason why the concepts of *common cycles* and *common serial correlation* can be used interchangeably.

Considering now that $C(1) = I - \sum_{j>0} C_j$, it follows that $C_0^* = C(1) - I$. Thus, if $\delta' C_0^* = 0$, then, $C(1)'\delta = \delta$. This condition implies that the *SCCF* vectors are the eigenvectors associated with the unit eigenvalues of $C(1)'$. Since the cointegration vectors are orthogonal to $C(1)$, they are the eigenvectors associated with the zero eigenvalues of $C(1)$. Thus, $\beta$ and $\delta$ if they exist, must be orthogonal, implying that if the cointegration rank is $r$ ($r < n$), then, there can be at most $s = n - r$ linearly independent *SCCF* vectors. This restriction is removed by the definition of *weak SCCF* (see §IV.2.3).

The testing procedure proposed by Vahid and Engle (1993) is very similar in nature to the

one proposed by Engle and Kozicki (1993) for stationary variables, described in §IV.2.1. The main distinction is that for cointegrated variables, the relevant past of the differences is defined not only by the lagged values of all the variables in the system, but also by the deviations form equilibria.

The procedure consists, therefore, in estimating the cointegration vectors and the error correction coefficients, and then, apply exactly the same procedure as Engle and Kozicki (1993) but including the deviations from equilibria (if any) as a component of the matrix containing the relevant past. That is, we need to find the smallest canonical correlations between $\Delta X_t$ and $[\hat{\beta} X_{t-1}, \Delta X_{t-1}, ..., \Delta X_{t-k+1}]$. The test statistic is given by eq. (IV.1).

### IV.2.3 Weak *SCCF* in cointegrated series, a natural extension

A natural extension of $SCCF$ in cointegrated series is to allow the possibility that the $SCCF$ vectors cancel the short-run dynamics, but are not related in a particular way with the long-run pattern of the series. That is, there could exist a linear combination of the differenced series that is an innovation with respect to the past, but only after adjusting for the equilibria deviations. This is the concept of *weak form reduced rank structures* (hereafter *WF*) introduced by Hecq et al. (2006).

Recall the error correction representation of the VAR model:

$$\Delta X_t = \Pi X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1} \Delta X_{t-k+1} + \epsilon_t, \tag{IV.4}$$

where $\Pi = -I_n + \Pi_1 + ... + \Pi_k$, $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$, and and $\Pi_i$ are the coefficient matrices of the VAR in levels. Given that the series are cointegrated, Hecq et al. (2006) distinguish two, not mutually exclusive, possibilities for the $SCCF$ vector ($\delta$):

$$\delta' \Phi_i = 0_{(s \times n)}, \ i = 1, .., k-1, \ \ and \tag{IV.5}$$

$$\delta' \Pi = \delta' \alpha \beta' = 0_{(s \times n)}. \tag{IV.6}$$

A $SCCF$ will be present in cointegrated series if both conditions eq. (IV.5) and eq. (IV.6) are fulfilled. The novelty of Hecq et al.'s (2006) proposal is for cointegrated series that satisfy eq. (IV.5) but not eq. (IV.6). This is the case of *WF*, which is a necessary but not sufficient condition for $SCCF$. The detailed discussion of estimation and testing procedures will be covered

later on, in §IV.2.5.

These new short-run common structures have, at least, three interesting implications over the model dynamics: i) Since the *WF* vectors are not required to be linearly independent of the cointegrating vectors, the number of *WF* vectors is not limited by the number cointegration relations. ii) As Hecq et al. (2006) show, in the cointegrated VAR(1) with $s > r$, condition eq. (IV.6) implies the existence of $s - r$ *SCCF* vectors. iii) Under the *WF* structure, the short and long-run dynamics of the series are unrelated, this contrasts with the *SCCF*, where the serial correlation patterns of $\Delta X_t$ and $\alpha\beta' X_t$ are the same.

### IV.2.4 Polynomial serial correlation common feature

Even after the generalization to consider WF, the requirements for the short-run comovements for accepting the existence of a common correlation feature can result too strong in some cases. The main drawback of Engle and Kozicki's (1993) concept of *SCCF* was noted by Ericsson (1993) in a comment on their article. Ericsson (1993), argues that a common correlation feature may exists in a multivariate time series, but it does not need to be contemporaneous as SCCF require. If the linear combination of the series that eliminates the feature requires some of them to be lagged, it does not seem reasonable to disregard the existence of a common feature as would do the *SCCF* test. This issue motivated Vahid and Engle (1997) to extend the concept of *SCCF* to *co-dependence*, which deems the possibility of not exactly synchronized cycles.

Cubadda and Hecq (2001), introduce the concept of *polynomial serial correlation common feature* (*PSCCF*) as an alternative measure of non-contemporaneous cyclical co-movements. The main motivation of the authors is that the co-dependence measure proposed by Vahid and Engle (1997) requires elaborate procedures for inference because the reduced rank regression technique in not applicable for estimating co-dependence vectors.

Cubadda and Hecq (2001) focus their attention on I(1) time series, for which cointegration relations may exist. The equilibrium correction representation of eq. (IV.4) can be rewritten as $[\Pi(1)L + \Phi(L)\Delta]X_t = \epsilon_t$, where $\Phi(L) = I_n - \sum_{i=1}^{k-1} \Phi_i L^i$. Additionally, as noted by Cubadda and Hecq (2001), the polynomial matrix $\Pi(L)$ can be expressed as: $\Pi(L) = \Delta - A(L)$, where

$$A(L) = \alpha\beta' L + \Phi^*(L)\Delta, \; and \;\; \Phi^*(L) = \Phi(L) - I_n. \tag{IV.7}$$

Ignoring deterministic components, we can write the stationary process $\Delta X_t$ in its Wald

representation form as $\Delta X_t = C(L)\epsilon_t$, where, $\sum_{j=1}^{\infty} j|C_j| < \infty$ and $C_0 = I$. Considering the VAR model $\Pi(L)X_t = \epsilon_t$, and the Wald representation, we obtain:

$$C(L)\Pi(L)X_t = \Delta X_t \tag{IV.8}$$

Cubadda and Hecq (2001) define the *PSCCF* as:

**Definition 1** *The series $\Delta X_t$ have s PSCCF of order m, noted as $PSCCF(m)$, if and only if there exist a $(n \times s)$ polynomial matrix $\delta(L) = \sum_{i=0}^{m} \delta_i L^i$ such that the matrix $\delta_0$ has full column rank and $\delta(L)'C(L) = \delta_0'$.*

The case of major interest for the authors is $m = 1$, so that $\delta(L) = \delta_0 + \delta_1 L$. In this case, the conditions for the existence of a *PSCCF(1)* can be easily derived. Using definition 1, pre-multiplying both sides of eq. (IV.8) by $\delta(L)'$, and using $\Pi(L) = \Delta - A(L)$, we get:

$$-\delta_0' A(L)X_t = \delta_1' \Delta X_{t-1}. \tag{IV.9}$$

Then, there will be a *PSCCF(1)* structure if and only if eq. (IV.9) holds. Plugging the eq. (IV.7) in condition eq. (IV.9), we get:

$$-\delta_0'(\alpha\beta' L + \Phi^*(L)\Delta)X_t = \delta_1' \Delta X_{t-1}. \tag{IV.10}$$

Now, for eq. (IV.10) to hold, the following two conditions must be satisfied:

a) $\delta_0'\alpha = 0$, and b) $\delta_0'\Phi_i = \begin{cases} -\delta_1' & if\ i = 1 \\ 0 & if\ i > 1 \end{cases}$.

These conditions imply that $\delta_0'$ have to belong to the left null space of all the *VEqM* coefficient matrices except $\Phi_1$. Thus, the VEqM can again be seen as a reduced rank regression, so that the estimation of $\delta(L)$ requires to solve one of the following — equivalent — canonical correlation problems:

$$CanCor\left\{\Delta X_t, \begin{pmatrix} \hat{\beta} X_{t-1} \\ \Delta X_{t-2} \\ \vdots \\ \Delta X_{t-k+1} \end{pmatrix} | d_t, \Delta X_{t-1}\right\}, \quad or, \quad CanCor\left\{\begin{pmatrix} \Delta X_t \\ \Delta X_{t-1} \end{pmatrix}, \begin{pmatrix} \hat{\beta}' X_{t-1} \\ \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-k} \end{pmatrix} | d_t\right\}, \tag{IV.11}$$

where $CanCor(Y, Z|W)$ stands for the canonical correlation between $Y$ and $Z$ given $W$, and $d_t$ contains all deterministic components in model. The elements of $\delta_0$ and $\delta_1$ are given by the eigenvectors associated with the $s$ smallest eigenvalues in the second expression.

As showed by Anderson (2003), the hypothesis for the existence of at least $s$ zero eigenvalues (i.e., $s$ $PSCCF(1)$ linear independent vectors) can be tested with the following test statistic:

$$lnQ(\lambda) = -T\sum_{i=1}^{s} ln(1-\hat{\lambda}_i),\ s = 1, .., n, \tag{IV.12}$$

where $\hat{\lambda}_i$ are the ordered smallest eigenvalues (squared canonical correlations) derived form eq. (IV.11). The asymptotic distribution of the statistic is $\chi^2_v$, with $v = s[n(k-2)+r]-s(n-s)$.[1]

### IV.2.5 A general unifying framework to look for stationary common structures

So far, we have described a number of possible stationary linear common structures and their associated estimation and testing procedures. All of them are related, and even some of them are necessary conditions for others. The objective of this section is to put all these concepts in order and analyze the possibility of dealing with all of them in the same model. Cubadda (2007) gives an unifying framework for deeming all the short-run comovement possibilities at the same time. His approach can be seen as an extension of either Cubadda and Hecq's (2001) $PSCCF$, or the Hecq et al.'s (2006) *weak form of reduced rank structure* (WF). In fact, it is a combination of both, and allows for both kinds of structures at the same time. This wider concept is called *weak form polynomial serial correlation* (hereafter WFP), and allows all the comovement structures considered along this section. Though we do not consider WPF neither in the simulation studies nor in the application, in the following lines we give a brief description of the concept.

Consider the VEqM model of eq. (IV.4). As defined by Cubadda (2007):

**Definition 2** [2] *A series $\Delta X_t$ has $s$ WFPs of order one if and only if there exists a $(n \times s)$ matrix $\delta_F$ with full column rank such that $\delta'_F\alpha \neq 0$, $\delta'_F\Phi_1 \neq 0$, and the VEqM in eq. (IV.4) can be rewritten as:*

$$\Delta X_t = \alpha\beta' X_{t-1} + \Phi_1\Delta X_{t-1} + \delta_{F\perp}\Psi'_F(\Delta X'_{t-2}, ..., \Delta X'_{t-k+1})' + \epsilon_t, \tag{IV.13}$$

*where $\delta_{F\perp}$ is the orthogonal complement of $\delta_F$; and $\Psi_F$ is a $(nk-2n)\times(n-s)$ full column*

[1] We are assuming that $k > 2$, since otherwise a trivial $SCCF$ will appear.
[2] Slight adaptation of the original version in Cubadda (2007).

*rank matrix, such that* $\delta_{F\perp}\Psi'_F = [\Phi_2, ..., \Phi_{k-1}]$.

Multiplying both sides of eq. (IV.13) by $\delta'_F$ we get: $[\delta'_F - \delta'_F(I_n + \alpha\beta' + \Phi_1)L + \delta'_F\Phi_1 L^2]X_t = \delta'_F\epsilon_t$. Thus, it results that a WFP structure of order one implies the existence of a second order polynomial matrix:

$$\delta_F(L) = \delta_F - (I_n + \beta\alpha' + \Phi'_1)\delta_F L + \Phi'_1\delta_F L^2, \quad \textit{such that} \tag{IV.14}$$

$$\delta_F(L)'X_t = \delta'_F\epsilon_t \tag{IV.15}$$

Note that, for simplifying the notation, we are making abstraction of deterministic components in the VEqM model. When they exist, they would have to be added to the right hand side of eq. (IV.15) because WFP do not require the cancellation of the deterministic structure.

In view of, eq. (IV.13) and similarly to Cubadda and Hecq (2001), inference for WFP can be carried out by solving a canonical correlation program:

$$CanCor\left\{\left(\Delta X'_{t-2}, \quad \ldots, \quad \Delta X'_{t-k+1}\right)', \left(1, \quad \Delta X'_{t-1}, \quad X'_{t-1}\hat{\beta}\right)'\right\}.$$

Inference for the other types of common features (*SCCF, PSCCF* and *WF*) is also carried out by canonical correlation problems for the appropriate data sets (see Tables 1 and 2 in Cubadda, 2007), and the test statistic is given by eq. (IV.12).

#### IV.2.5.1 Simultaneous testing

Up to now, via canonical correlation problems for the appropriate data sets, we are able to test for the existence of the different kinds of common features (*SCCF, PSCCF, WF* and *WFP*). Cubadda (2007) proposes a specification of the model that allows to test all the different possibilities simultaneously in the same model. This has the evident advantage of allowing the detection of more than one kind of common feature at the same time. The general model has the following form:

$$v_t = \delta_\perp\Psi' w_{t-1} + \eta_t \tag{IV.16}$$

where $v_t = (\Delta X'_t, X'_{t-1}\beta, \Delta X'_{t-1})'$ is $(2n + r) \times 1$; $w_{t-1} = (X'_{t-1}\beta, \Delta X'_{t-1}, ..., \Delta X'_{t-k+1})'$; $\eta_t = (\epsilon'_t, 0_{1\times(r+n)})'$; $\delta$ is a $(2n + r) \times s$ matrix with $s < n$; and $\Psi$ is a $(r + kn - n) \times (2n + r - s)$

matrix such that:

$$\delta_{\perp}\Psi' = \begin{pmatrix} (\alpha, \Phi_1) & (\Phi_2, ..., \Phi_{k-1}) \\ I_{r+n} & 0_{(r+n)\times(kn-2n)} \end{pmatrix}_{(r+2n)\times[r+n(k-1)]}$$

Note that eq. (IV.16) is a compacted representation of eq. (IV.13), and the WFP vectors can be tested by: $CanCor\{v_t, w_{t-1}\}$. Similarly, all the other forms of common features can be estimated by imposing the adequate restrictions in eq. (IV.16). Letting $\delta = H\theta$, where $H$ is a $(2n + r) \times g$ 'restriction matrix' of full column rank, and $\theta$ is the $g \times s$ matrix of restricted parameters to be estimated, the tests are preformed by:

$$CanCor\{Hv_t, w_{t-1}\}. \tag{IV.17}$$

Table 3 in Cubadda (2007) gives the specific form for the matrix $H$ associated with each type of common feature (*SCCF, PSCCF* and *WF*). The LR test statistic for the null $\delta = H\theta$ is given by:

$$Q = T\sum_{i=1}^{s} ln(\frac{1-\hat{\lambda}_i}{1-\hat{\omega}_i}),\ s = 1, ...n, \tag{IV.18}$$

where $\hat{\lambda}_i$ and $\hat{\omega}_i$ are the $i$th smallest eigenvalues from $CanCor\{v_t, w_{t-1}\}$ and eq. (IV.17) respectively. The test statistic has a $\chi^2$ distribution with $s(2n + r - g)$ degrees of freedom.

Although Cubadda (2007) stated it for contrasting stationary commonalities, eq. (IV.16) also allows to perform cointegration tests. It is therefore a very general expression that allows not only the individual contrast of all the common features considered until now, but also has the potential for simultaneous contrasts. Cubadda (2007) deems two separate cases of simultaneous contrasts; coexistence of PSCCF, and WF and the existence of nested common feature structures. The author gives the procedures and the test statistics for each case.

To close this section, it is worth to make a comment about the implications of the more comprehensive short-run comovement structure proposed by Cubadda (2007) (WFP) on the cycles of the variables' levels. In §IV.2.2, we stated that the existence of SCCF structures in the first differences of I(1) variables was a necessary and sufficient condition for the existence of common cycles — in the sense of Beveridge and Nelson (1981). This parallelism makes SCCF to be directly interpretable form an economic theory viewpoint. However, as Cubadda (2007) shows, WFP structures are not so tightly related with the existence of common cycles. Indeed,

the combination of the series' cycles trough the polynomial $\delta_F(L)'$ results in a VMA(1) process, instead of white noise as was the case for SCCF structures.

## IV.3 Relationship between Common Features and Common Factors

In this section we review the relationship between the common features studied in §IV.2 and DFM. We split the analysis into two main sections, while in §IV.3.2 we use a VAR model as the original DGP and derive the equivalent DFM, in §IV.3.3, the DGP is a DFM and we derive the corresponding VARMA. We split §IV.3.2 into three subsections in which the DGP has only cointegration restrictions, cointegration and SCCF, and cointegration with WF. Similarly, we split §IV.3.3 into two subsections that include only I(1) factors, and I(1) plus I(0) factors. In §IV.3.1 we include a summary of the results, so that the reader not interested in the details can focus only in that subsection and then jump to §IV.4.

The main contribution of this section is the analysis of the connection between models with common stationary features and models with $I(0)$ common factors. For the best of our knowledge, this analysis does not exist in the literature. The connection between models with common trends and common $I(1)$ factors already exists in the literature, and we give a review of it. The fact that we summarize the connections between the two types of models in a single framework, also constitutes a contribution.

The interest of this section in the context of the thesis is that it will be necessary to understand the simulation results of §IV.5, in which both VAR models and DFM are used as data generating processes.

For readers not familiar with the DFM literature, appendix C.1 includes a review of this literature.

### IV.3.1 Summary

i. **DGP is a VAR:**

(a) **Only cointegration ($r$ restrictions):** the equivalent DFM for $X_t$ has $(n-r)$ common $I(1)$ factors and stationary idiosyncrasies. If the DGP has non-pervasive common trends the DFM's idiosyncrasies will be $I(1)$ and the 'difference recumulating' proce-

dure proposed by Bai and Ng (2004) should be applied.

(b) **Cointegration and SCCF:** the equivalent DFM for $X_t$ has $(n-r)$ $I(1)$, and $(n-s)$ $I(0)$ factors and no idiosyncrasies.

(c) **Cointegration and WF:** from the MA representation in levels, we observe only the $n-r$ common $I(1)$ factors, common stationary factors do not appear. From the VEqM representation, a DFM with $r+(n-s)$ stationary factors can be derived, and we cannot distinguish between trends and cycles. Which of the two representations will show up in empirical applications is an empirical matter.

ii. **DGP is a DFM**

(a) **Only $I(1)$ factors:** as showed by Peña and Poncela (2006a), the reduced form is a VARMA with equilibrium correction. Although in the DGP there are no $I(0)$ factors, the corresponding VARMA will have WF if the idiosyncrasies of the DFM are not serially correlated.

(b) **Only $I(0)$ factors:** if idiosyncrasies are not serially correlated, the reduced form is a *VMA* with *SCCF*. If idiosyncrasies are serially correlated, there will be no common cycles restrictions as defined by Cubadda (2007). There are, however, *scalar component models* (SCM) structures as defined by Tiao and Tsay (1989).

(c) **$I(1)$ and $I(0)$ factors:** the reduced form is a VARMA with equilibrium correction, but no common cycles restrictions, even when idiosyncrasies are not serially correlated. Again, there are SCM structures.

## IV.3.2 Original DGP is a VAR

### IV.3.2.1 Restrictions in the DGP: only cointegration

In this subsection we start from a VEqM and derive the equivalent DFM. Consider the VEqM form of a cointegrated VAR model (ignoring deterministic components):

$$\Delta X_t = \alpha\beta' X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1}\Delta X_{t-k+1} + \epsilon_t, \tag{IV.19}$$

where $X_t$ is a $n \times 1$ vector; $\alpha$ and $\beta$ are $n \times r$ full column rank matrices; $r$ is the cointegration rank; $\alpha\beta' = -I_n + \Pi_1 + ... + \Pi_k$, $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$; $\Pi = -\Pi(1)$; matrices $\Pi_i$ come from the VAR representation; and $\epsilon_t \sim N(0, I_n)$.

From the Granger Representation theorem (see e.g Engle and Granger (1987), or Johansen (1991)) the VEqM in eq. (IV.19) can be written as:

$$X_t = X_0 + C(1)\sum_{i=1}^{t}\epsilon_t + C^*(L)\epsilon_t, \tag{IV.20}$$

where $C(1) = \beta_\perp(\alpha'_\perp\Psi\beta_\perp)^{-1}\alpha'_\perp$, is $n \times n$ but has rank $n - r$; $\Delta X_t = C(L)(\mu + \epsilon_t)$; $C(L) = C(1) + (1 - L)C^*(L)$; $\Psi(L) = (1 - L)^{-1}[\Pi(L) - \Pi(1)L]$; and $\Psi = \Psi(1)$.

The cointegrated process in eq. (IV.20) allows an alternative representation in terms of a reduced number of common random walks plus a stationary component (see, *inter alia*, Stock and Watson (1988), Gonzalo and Granger (1995), or Escribano and Peña (1994)). Here we follow the proof in Escribano and Peña (1994).

Defining the $(n - r) \times n$ matrix $H = (\alpha'_\perp\Psi\beta_\perp)^{-1}\alpha'_\perp$, so that $C(1) = \beta_\perp H$; the $n \times 1$ vector $e_t = C^*(L)\epsilon_t$; and the $(n - r) \times 1$ vector $v_t = H\epsilon_t$; eq. (IV.20) becomes:

$$\begin{aligned} X_t &= \beta_\perp\tau_t + e_t \\ \tau_t &= \tau_{t-1} + v_t \end{aligned} \tag{IV.21}$$

Note that in eq. (IV.21), the vector of common stochastic tends $(\tau_t)$ has dimension $n - r$ instead of $n$ as in eq. (IV.20). Given that $\tau_t$ contains all system's unit roots, idiosyncrasies $e_t$ are stationary. This transformation to a model in which the unit roots in the system depend only on $n - r$ random walks was possibly only because of the reduced rank condition of matrix $C(1)$. In absence of this condition, $C(1)$ could not had been expressed as the product of two matrices with reduced column and row dimensions, respectively.

Consider now the Dynamic Factor Model

$$\begin{aligned} X_t &= \Lambda F_t + u_t \\ (I - L)F_t &= G(L)a_t \\ (1 - \rho_i L)u_{it} &= D_i(L)s_{it}, \end{aligned} \tag{IV.22}$$

where the dimension of $F_t$ is the number of common factors. If G(1) has full rank, all factors are I(1), otherwise some of them are I(0). If $\rho_i < 1$ for all $i$, idiosyncrasies are $I(0)$. Assuming

that there are $n-r$ I(1) factors and $k-n+r$ stationary ones, we can write their models as:

$$(I-L)\tau_t = G_1(L)a_{1t},$$
$$w_t = G_2(L)a_{2t},$$

with $[\tau_t', w_t']' = F_t'$, so that $X_t = \Lambda_1\tau_t + \Lambda_2 w_t + u_t$. Then, $X_t$ can be written as:

$$\begin{aligned} X_t &= \Lambda_1\tau_t + e_t, \\ \tau_t &= \tau_{t-1} + v_t, \end{aligned} \tag{IV.23}$$

where $e_t = \Lambda_2 w_t + u_t$ and $v_t = G_1(L)a_{1t}$. Clearly, eq. (IV.21) is a special case of eq. (IV.23) in which matrix $\Lambda_1$ and vectors $e_t$ and $v_t$ satisfy certain conditions.

#### IV.3.2.2 An implication for Dynamic Factor analysis

As showed by Escribano and Peña (1994), the common trends representation of the VEqM in eq. (IV.21) is a particular case of eq. (IV.22) and eq. (IV.23), in which the factor loading matrix $\beta_\perp$ is the orthogonal complement of the cointegration matrix $\beta$. This feature ensures that idiosyncrasies are stationary, as all the unit roots of the system are included in the common factor. Indeed, the idiosyncratic error $e_t$ in eq. (IV.21) is equal to $C^*(L)\epsilon_t$, which is a stationary process.

Consider now a case where the number of cointegrating relationships is small relative to the number of series ($r << n$), or even being large, there is a subset of series not sharing the trend with any other in $X_t$ (these are the cases found by Espasa and Mayo-Burgos (2013) when analyzing the CPI of US, UK, and the Euro Area). Under these frameworks, some of the system's unit roots will be idiosyncratic, or common just to a reduced number of series, so that matrix $\beta_\perp$ will have rows full of zeros, except for some few non zero elements.

To get a visual example, consider an extreme case in which vector $X_t$ has dimension 100, the first 10 series share the trend, but the other 90 show their own long run dynamics. Without loss of generality, we can normalize matrix $\beta$ such that $\beta^* = [\beta^* : I_n]$ (see Clements and Hendry (1995)), thus, assuming for visual simplicity that all the elements of $\beta^*$ are equal to -1, $\beta$ will be:

$$\beta = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{100\times 9},$$

so that:

$$\beta_\perp = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{100\times 91}.$$

This factor loading structure implies that 90 out of the 91 'common trends' are in fact idiosyncratic, not common. Thus, model eq. (IV.21) is not a proper factor model since almost none of the factors are truly common.

To avoid this issue, the model could be reformulated so that the loading matrix has only one column and idiosyncrasies are non stationary. Then, when there are trends that affect only a reduced subset of series, or $r << n$, idiosyncrasies will be non stationary and the 'differencing recumulating' procedure proposed by Bai and Ng (2004) should be applied.

#### IV.3.2.3 Restrictions in the DGP: cointegration and SCCF

This is the case analyzed above but augmented by the presence of common cycles. Recall the VMA model for $X_t$ of eq. (IV.20):

$$X_t = X_0 + C(1)\sum_{i=1}^{t}\epsilon_t + C^*(L)\epsilon_t. \tag{IV.24}$$

The presence of common trends depends on the rank of matrix $C(1)$, similarly, the different types of stationary common features are associated with the rank of matrices $C_i^*$.

The connection between SCCF and stationary common factors was analyzed by Vahid and Engle (1993). The authors show that "*if the left null spaces of a set of $n \times n$ matrices overlap, and the dimension of the intersection is $s$, then they all can be written as the product of two matrices with dimensions $n \times (n-s)$ and $(n-s) \times n$ with the left matrix having full column rank and being the same for all of them*". Thus, if there are $s$ linearly independent *SCCF* vectors, we can write:

$$C_i^* = \delta_\perp \tilde{C}_i^*, \quad \text{for all} \quad i \geq 0$$

with $\delta_\perp$ having dimensions $n \times (n-s)$ and being a base of the intersection of the left null spaces of matrices $C_i^*$. Hence, the Beverdige-Nelson cycle $C^*(L)\epsilon_t$, can be written as:

$$c_t = C^*(L)\epsilon_t = \delta_\perp \tilde{C}^*(L)\epsilon_t \equiv \delta_\perp \tilde{c}_t. \tag{IV.25}$$

In the same fashion as the common trends representation, in eq. (IV.25) we have expressed the cycles of the $n$ series as linear combinations of $n-s$ common cycles. Then, the common trends-common cycles representation of $X_t$ would be: [3]

$$\begin{aligned} X_t &= \beta_\perp \tau_t + \delta_\perp \tilde{c}_t \\ \tau_t &= \tau_{t-1} + v_t \\ \tilde{c}_t &= \tilde{C}^*(L)\epsilon_t. \end{aligned} \tag{IV.26}$$

Note the difference between eq. (IV.26) and eq. (IV.21): while in eq. (IV.21) we have $n$ idiosyncratic components ($e_t = C^*(L)\epsilon_t$), in eq. (IV.26) there are no idiosyncrasies since the stationary part of $X_t$ also presents commonalities.

Then, if the factors are pervasive, fitting a DFM to $X_t$ or $\Delta X_t$ will easily lead to discover the common trends and common cycles. In presence of non-pervasive 'common trends' the procedure in Bai and Ng (2004) should be applied.

[3] Recall that $\beta$ and $\delta$ contain the eigenvectors of $C(1)'$ associated with zero and unit eigenvalues respectively, so that their columns are orthogonal between each other.

#### IV.3.2.4 Restrictions in the DGP: cointegration and WF

From definition 3 in Cubadda (2007), the vector $\Delta X_t$ have $s$ WFs if and only if there exists an $n \times s$ matrix $\delta_W$ with full column rank such that $\delta'_W \alpha \neq 0$, and the VEqM in eq. (IV.19) can be rewritten as the following partial RRR model (to ease notation we are making abstraction of deterministic components):

$$\Delta X_t = \alpha\beta' X_{t-1} + \delta_{W\perp}\Psi'_W(\Delta X'_{t-1}, ..., \Delta X'_{t-k+1})' + \epsilon_t, \tag{IV.27}$$

with $\delta'_W \delta_{W\perp} = 0$; and $\Psi_W$ a $(n \times k - n) \times (n - s)$ matrix with full column rank.

We now derive the MA representation of previous model. Cubadda (2007) shows that the definition of WF is equivalent to the existence of a first order polynomial matrix $\delta(L)$ such that $\delta'(L)X_t$ is an innovation, and:

$$\delta_W(L) = \delta_W - (\beta\alpha' + I_n)\delta_W L. \tag{IV.28}$$

As an interesting result, the author shows that:

$$\delta'_W(L)C^*(L)\epsilon_t = \delta'_W(I_n - C(1))\epsilon_t, \tag{IV.29}$$

so that $\delta'_W(L)C^*(L)\epsilon_t$ is also an innovation.

Note now that coefficients matrices $\Phi_i$ of eq. (IV.19) are $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$, thus, eq. (IV.27) implies:

$$\delta'_W \Pi_i = 0 \text{ for all } 2 \leq i \leq k. \tag{IV.30}$$

Then, using $\alpha\beta' = -\Pi(1)$, the polynomial in eq. (IV.28) can be simplified to:

$$\delta_W(L) = \delta_W - \Pi'_1 \delta_W L, \tag{IV.31}$$

with $\Pi_1$ being the coefficients' matrix associated with the first lag of the VAR in levels. Using eq. (IV.29), eq. (IV.31), and using the method of indeterminate coefficients we get:

$$\begin{aligned} C_0^* &= I_n - C(1), \\ \delta_W'(C_1^* - \Pi_1 C_0^*) &= 0, \\ &\vdots \\ \delta_W'(C_k^* - \Pi_1 C_{k-1}^*) &= 0. \end{aligned} \tag{IV.32}$$

If the WF structure implied a stationary common factor, we should expect a reduced rank condition on some of the $C_i^*$, for some $i > 0$. In such a case, we would be able to write part of the series' stationary dynamics as a function of a process with dimension lower than $N$. Thus, we would need $\delta_W' C_i^* = 0$, for *some* $i$. Note that if $\delta_W' C_i^* = 0$ for some $i$, in order to to satisfy eq. (IV.32), we would also need $\delta_W' \Pi_1 = 0$. But if $\delta_W' \Pi_1 = 0$, eq. (IV.32) requires $\delta_W' C_j^* = 0$, for *all* $j > 0$, which implies a $SCCF$ structure.

Therefore, $\delta_W' C_i^* = 0$ for *some* $i$, implies $\delta_W' C_i^* = 0$ for *all* $i$, which lead us to the $SCCF$ case. Therefore, in the pure WF case the vector $\delta_W'$ cannot cancel any short run matrix of the MA representation, so that in this representation stationary common factors do not show up directly.

Focus now on the VEqM representation of eq. (IV.27). In that expression, it can be easily seen that $\Delta X_t$ can be written as a two factor model:

$$\Delta X_t = \alpha F_{1t} + \delta_{W\perp} F_{2t} + \epsilon_t, \tag{IV.33}$$

where $\alpha$ is $n \times r$; $\delta_{W\perp}$ is $n \times (n-s)$; $F_{1t} = \beta' X_{t-1}$; and $F_{2t} = \Psi_W'(\Delta X_{t-1}', ..., \Delta X_{t-k+1}')'$. Therefore, from this representation we will obtain a factor model with $r + (n-s)$ factors. Note, however, that we will not be able to distinguish between common trends and common cycles, as the integrated factors will appear as I(1).

In sum, we have argued that a VAR with cointegration and WF restrictions can be seen either as a DFM with $n - r$ common trends, or as a DFM with $r + (n - s)$ common factors. Which of the two representations will emerge in real data problems, depends on which factor structure represents a larger proportion of the variance. It is, therefore, an empirical matter. Note that if the cointegration relationships and the 'common cycles' are pervasive, the error component, $\epsilon_t$, in eq. (IV.33) is white noise. This is not the case for the DFM derived from the MA representation, in which the error contains all the stationary structure of the series. Therefore, under pervasive cointegration relationships and 'common cycles', the DFM that would emerge

from the data could be expected to have $r + (n - s)$ common factors.

### IV.3.3 Original DGP is a DFM

#### IV.3.3.1 Only $I(1)$ factors

We now start from a DFM and derive the equivalent VARMA model. Assume first that the factors are $I(1)$ so that $G(1)$ in eq. (IV.22) has full rank. Then;

$$F_t = F_{t-1} + G(L)a_t \tag{IV.34}$$

From the first line in eq. (IV.22) we can write:

$$F_t = \Lambda^+(X_t - u_t), \tag{IV.35}$$

with $\Lambda^+ = (\Lambda'\Lambda)^{-1}\Lambda'$. Hence, using eq. (IV.22), eq. (IV.34) and eq. (IV.35) we get:

$$X_t = \Lambda\Lambda^+(X_{t-1} - u_{t-1}) + \Lambda G(L)a_t + u_t,$$

so that;

$$(1 - L)X_t = (\Lambda\Lambda^+ - I)X_{t-1} + \Lambda G(L)a_t + (I - \Lambda\Lambda^+ L)u_t. \tag{IV.36}$$

Equation (IV.36) is the VEqM implied by the DFM derived by Peña and Poncela (2006a). As noted by the authors, $\Lambda\Lambda^+$ has rank $n - r$ and since it is an idempotent matrix, it has all its eigenvalues equal to one or zero. Thus, $(\Lambda\Lambda^+ - I)$ has rank $r$, and $(\Lambda\Lambda^+ - I)X_{t-1}$ can be seen as the equilibrium correction term.

Note also that, although there are not $I(0)$ common factors in eq. (IV.34), if $u_t$ is white noise, the linear combination $\Lambda_\perp(1 - L)X_t$ is white noise plus $\Lambda_\perp(\Lambda\Lambda^+ - I)X_{t-1}$, so that there will be $(n - r)$ WF restrictions.

Finally, recall that the canonical correlation tests discussed in §IV.2 assume a VAR structure for $X_t$, however, when the data comes from a DFM we will also have a MA component, as eq. (IV.36) indicates.

#### IV.3.3.2 $I(1)$ and $I(0)$ factors

Assume a first DFM with only $(n - s)$ $I(0)$ factors:

$$\begin{aligned} X_t &= \Lambda F_t + u_t \\ F_t &= G(L)a_t \\ (1-\rho_i L)u_{it} &= D_i(L)s_{it} \end{aligned} \quad .$$

Then, $X_t = \Lambda G(L)a_t + u_t$. If $u_t$ is white noise, $X_t$ follows a *VMA(k)* model — it can be easily seen that the autocorrelation function of $X_t$ is zero except for the first $k$ values, being $k$ the order of $G(L)$ — and there exist a linear combination $\Lambda_\perp X_t$ that is white noise, with $\Lambda_\perp$ being an $s \times n$ matrix. Hence, there are $s$ *SCCF* restrictions and $(n-s)$ common cycles. Note, however, that when the DGP is a DFM we will have a VMA structure instead of a VAR one.

If $u_t$ is not white noise but $\rho_i = 0\ \forall i$, $X_t$ still follows a VMA process but $\Lambda_\perp X_t$ is not any more white noise and there will be no *SCCF* restrictions.

If $\rho_i \neq 0$ $X_t$ will have a VARMA representation with no *SCCF* restrictions, but there would be a *scalar component models* (SCM) structure as defined by Tiao and Tsay (1989).

Consider now a more general case where there are both, $I(0)$ and $I(1)$ factors:

$$\begin{aligned} X_t &= \Lambda F_t + u_t \\ (I-L)F_t &= G(L)a_t \\ (1-\rho_i L)u_{it} &= D_i(L)s_{it}, \end{aligned}$$

assuming that $G(1)$ does not have full rank we have some $I(0)$ and some $I(1)$ factors. Assuming further that the models for the factors can be written as:

$$\begin{aligned} (I-L)F_{1t} &= G_1(L)a_{1t} \\ F_{2t} &= G_2(L)a_{2t}, \end{aligned}$$

we get $X_t = \Lambda_1 F_{1t} + \Lambda_2 F_{2t} + u_t$. Following the same arguments we used above to derive eq. (IV.36) we get:

$$(1-L)X_t = (\Lambda_1\Lambda_1^+ - I_n)X_{t-1} + (I_n - \Lambda_1\Lambda_1^+ L)[u_t + \Lambda_2 G_2(L)a_{2t}] + \Lambda_1 G_1(L)a_{1t} \qquad \text{(IV.37)}$$

As in eq. (IV.36), matrix $(\Lambda_1\Lambda_1^+ - I_n)$ plays the role of $\alpha\beta'$ and we have an error correction model. A striking difference with respect to eq. (IV.36) is that now, even if $u_t$ were white noise, we will not have stationary common features. Again, a *SCM* also appears in this case.

## IV.4 On the properties of the pairwise approach applied to common cycles

The pairwise procedure applied to common cycles consists of performing common cycles tests between all the $N(N-1)/2$ pairs that exist in a set of $N$ series, and then, looking for subsets in which all the pairs showed a common cycle.

In this section we study the properties of this procedure, focusing on three issues. First, for the pairwise strategy to be sensible, the common cycles definitions to be used must be transitive. In §IV.4.1 we study the transitiveness of the different definitions and show that SCCF is the only transitive one. However, in §IV.4.2 we propose a slight modification of WF and PSCCF tests to make these definitions also transitive.

Second, in §IV.4.3 we analyze the asymptotic properties of the pairwise approach. As in chapter II, we are interested in the probability of including a high proportion of the correct series in the estimated subset, and in the probability of including wrong series ('potency' and 'gauge'). The dependency of the tests inside the 'single-cycle' subset is a critical aspect, and we show that under some conditions they can be independent. This independence makes the probability of discovering the true 'single-cycle' subset a fast decreasing function of the subset's size, what is an undesirable property. However, we argue and show by Monte Carlo, that a relaxation procedure — similar to that defined in §III.3 — virtually solves this problem with almost zero cost, asymptotically.

The third issue that we consider in this section, is the relative performance of the pairwise approach with respect to a full model strategy, when the number of series is small and the latter approach feasible. In §IV.4.4 we carry out this comparison by Mote Carlo and conclude that, when common cycles are pairwise detectable, nothing is lost for proceeding by pairs with respect to the full model approach. On the contrary, significant power gains for discovering the true number of common cycles may emerge from the pairwise procedure.

### IV.4.1 Transitivity

This subsection is devoted to analyze the transitiveness of the alternative definitions of common cyclical features. As aforementioned, we show that although the original $SCCF$ is the only definition that generally satisfies the transitivity property, a slight generalization of both $WF$ and $PSCCF$ makes them to become transitive.

#### IV.4.1.1 SCCF

To see that $SCCF$ is transitive let $X_1$ and $X_2$ be two $I(1)$ series that share a $SCCF$, and write:

$$\begin{aligned}\Delta X_{1t} &= c_1^1 + \varphi_1^1 CC_{t-1}^1 + \epsilon_{1,t}^1,\\ \Delta X_{2t} &= c_2^1 + \varphi_2^1 CC_{t-1}^1 + \epsilon_{2,t}^1,\end{aligned} \tag{IV.38}$$

where $CC_{t-1}^1 = \psi'[\Delta X'_{t-1}, ..., \Delta X'_{t-k+1}]'$; $[\varphi_1^1, \varphi_2^1]' = \delta_\perp$; $\epsilon_{i,t}^1 \sim WN$; and $\Delta X_t = [\Delta X'_{1t}, \Delta X'_{2t}]'$.

If $X_1$ and $X_3$ also share a $SCCF$, then:

$$\begin{aligned}\Delta X_{1t} &= c_1^2 + \varphi_1^2 CC_{t-1}^2 + \epsilon_{1,t}^2,\\ \Delta X_{3t} &= c_2^2 + \varphi_2^2 CC_{t-1}^2 + \epsilon_{2,t}^2.\end{aligned} \tag{IV.39}$$

Equalizing the first line of eq. (IV.38) and eq. (IV.39), solving for $CC_{t-1}^2$, and plugging the result in the second line of eq. (IV.39) we get:

$$\begin{aligned}\Delta X_{3t} = c_2^2 + \frac{\varphi_2^2}{\varphi_1^2}[(c_1^1 - c_1^2) + \varphi_1^1 CC_{t-1}^1 + (\epsilon_{1,t}^1 - \epsilon_{1,t}^2)] + \epsilon_{2,t}^2 \;\Rightarrow\\ \Delta X_{3t} = c_3 + \varphi_3 CC_{t-1}^1 + \epsilon_{3,t},\end{aligned} \tag{IV.40}$$

where $c_3 = c_2^2 + \frac{\varphi_2^2}{\varphi_1^2}(c_1^1 - c_1^2)$; $\varphi_3 = \frac{\varphi_2^2 \varphi_1^1}{\varphi_1^2}$; and $\epsilon_{3t} = \frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{1,t}^1 - \epsilon_{1,t}^2) + \epsilon_{2t}^2$. Since $\epsilon_{3t} \sim WN$, $X_3$ has the same $SCCF$ as $X_1$ and $X_2$.

Another way to see the transitivity of the $SCCF$ is to notice $CC_{t-1}^2$ can be written as a linear function of $CC_{t-1}^1$ plus a constant and a white noise. For the other cyclical common features defined in §IV.2 this argument works only under special conditions, as we show below.

#### IV.4.1.2 PSCCF

Consider the case of the $PSCCF$ as defined by Cubadda and Hecq (2001) and described in §IV.2.4. Equation (IV.38) and eq. (IV.39) become now:

$$\begin{aligned}\Delta X_{1t} &= c_1^1 + \phi_{11}^1 \Delta X_{1,t-1} + \phi_{12}^1 \Delta X_{2,t-1} + \varphi_1^1 CC_{t-2}^1 + \epsilon_{1,t}^1,\\ \Delta X_{2t} &= c_2^1 + \phi_{21}^1 \Delta X_{1,t-1} + \phi_{22}^1 \Delta X_{2,t-1} + \varphi_2^1 CC_{t-2}^1 + \epsilon_{2,t}^1.\end{aligned} \tag{IV.41}$$

$$\begin{aligned}\Delta X_{1t} &= c_1^2 + \phi_{11}^2 \Delta X_{1,t-1} + \phi_{12}^2 \Delta X_{3,t-1} + \varphi_1^2 CC_{t-2}^2 + \epsilon_{1,t}^2,\\ \Delta X_{3t} &= c_2^2 + \phi_{21}^2 \Delta X_{1,t-1} + \phi_{22}^2 \Delta X_{3,t-1} + \varphi_2^2 CC_{t-2}^2 + \epsilon_{2,t}^2.\end{aligned} \tag{IV.42}$$

Equalizing the first lines of eq. (IV.41) and eq. (IV.42), solving for $CC^2_{t-2}$, and plugging the result in the second line of eq. (IV.42) we get:

$$\Delta X_{3t} = c_2^2+\frac{\varphi_2^2}{\varphi_1^2}(c_1^1-c_1^2)+[\phi_{21}^2+\frac{\varphi_2^2}{\varphi_1^2}(\phi_{11}^1-\phi_{11}^2)]\Delta X_{1,t-1}+(\phi_{22}^2+\frac{\varphi_2^2}{\varphi_1^2}\phi_{12}^2)\Delta X_{3,t-1}+\frac{\varphi_2^2}{\varphi_1^2}\phi_{12}^1\Delta X_{2,t-1}+ \frac{\varphi_2^2}{\varphi_1^2}\varphi_1^1 CC_{t-2}^1+\frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{1,t}^1-\epsilon_{1,t}^2)+\epsilon_{2,t}^2 \Rightarrow$$

$$\Delta X_{3t} = c_2^3+[\phi_{21}^2+\frac{\varphi_2^2}{\varphi_1^2}(\phi_{11}^1-\phi_{11}^2)]\Delta X_{1,t-1}+\phi_{22}^3\Delta X_{3,t-1}+\phi_{21}^3\Delta X_{2,t-1}+\varphi_2^3 CC_{t-2}^1+\epsilon_{2,t}^3, \quad \text{(IV.43)}$$

where $c_2^3 = c_2^2 + \frac{\varphi_2^2}{\varphi_1^2}(c_1^1 - c_1^2)$; $\phi_{22}^3 = (\phi_{22}^2 + \frac{\varphi_2^2}{\varphi_1^2}\phi_{12}^2)$; $\phi_{21}^3 = \frac{\varphi_2^2}{\varphi_1^2}\phi_{12}^1$; $\varphi_2^3 = \frac{\varphi_2^2}{\varphi_1^2}\varphi_1^1$; and $\epsilon_{2,t}^3 = \frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{1,t}^1 - \epsilon_{1,t}^2) + \epsilon_{2,t}^2$.

Thus, from eq. (IV.41) and eq. (IV.42) one cannot deduce, in general, that $\Delta X_{2t}$ and $\Delta X_{3t}$ share the same $PSCCF$. There are two conditions under which the $PSCCF$ would be transitive. The first one is:

$$\phi_{21}^2 + \frac{\varphi_2^2}{\varphi_1^2}(\phi_{11}^1 - \phi_{11}^2) = 0 \Leftrightarrow \phi_{21}^2\varphi_1^2 = \varphi_2^2(\phi_{11}^1 - \phi_{11}^2), \quad \text{(IV.44)}$$

that will be satisfied only by chance. The second possibility would be that $\Delta X_{1,t-1}$ can be expressed as a linear combination of $\Delta X_{2,t-1}$, $\Delta X_{3,t-1}$ and $CC_{t-2}^1$ plus white noise. In such a case, $\Delta X_{3t}$ in eq. (IV.43) could be written as a linear combination of $\Delta X_{3,t-1}$, $\Delta X_{2,t-1}$, and $CC_{t-2}^1$ plus a white noise. Therefore, we need:

$$\Delta X_{1,t-1} = \gamma_1\Delta X_{2,t-1} + \gamma_2\Delta X_{3,t-1} + \lambda CC_{t-2}^1 + v_t,$$

hence,

$$\frac{1}{\lambda}\begin{bmatrix}1, & -\gamma_1, & -\gamma_2\end{bmatrix}\Delta X_t = CC_{t-2}^1 + \frac{1}{\lambda}v_t, \quad \text{(IV.45)}$$

where $\Delta X_t = [\Delta X_{1t}', \Delta X_{2t}', \Delta X_{3t}']'$. Recall now that:

$$CC_{t-2}^1 = \psi'[\Delta X_{1,t-2}', \Delta X_{2,t-2}', ..., \Delta X_{1,t-k}', \Delta X_{2,t-k+1}']',$$

where $\psi$ is a $2(k-1)\times 1$ vector. Thus, there exist a vector $\psi_\perp$ such that $\psi_\perp\psi' = 0$. Multiplying both sides of eq. (IV.45) by $\psi_\perp$ we can find a linear combination of $\Delta X_t$ that is white noise, so

that we would be in the $SCCF$ case. Therefore, we conclude that the 'pure' $PSCCF$ is not transitive when estimated in a paiewise fashion.

#### IV.4.1.3 WF

We consider now the *weak form of serial correlation common feature* (WF) as defined by Hecq et al. (2006), and described §IV.2.3. If $X_1$ and $X_2$ are two $I(1)$ series that share a $WF$, we can write:

$$\begin{aligned}\Delta X_{1t} &= c_1^1 + \alpha_1^1 CR_{t-1}^1 + \varphi_1^1 CC_{t-1}^1 + \epsilon_{1,t}^1,\\ \Delta X_{2t} &= c_2^1 + \alpha_2^1 CR_{t-1}^1 + \varphi_2^1 CC_{t-1}^1 + \epsilon_{2,t}^1,\end{aligned} \tag{IV.46}$$

where $CR$ stands for cointegration relationship. If $X_1$ and $X_3$ also share a $WF$ then:

$$\begin{aligned}\Delta X_{1t} &= c_1^2 + \alpha_1^2 CR_{t-1}^2 + \varphi_1^2 CC_{t-1}^2 + \epsilon_{1,t}^2,\\ \Delta X_{3t} &= c_2^2 + \alpha_1^2 CR_{t-1}^2 + \varphi_2^2 CC_{t-1}^2 + \epsilon_{2,t}^2.\end{aligned} \tag{IV.47}$$

Equalizing the first lines of eq. (IV.46) and eq. (IV.47), solving for $CC_{t-1}^2$, and plugging the result in the second line of eq. (IV.47) we get:

$$\Delta X_{3t} = \frac{\varphi_2^2}{\varphi_1^2}(c_1^1 - c_1^2) + c_2^2 + (\alpha_2^2 - \frac{\alpha_1^2\varphi_2^2}{\varphi_1^2})CR_{t-1}^2 + \frac{\alpha_1^1\varphi_2^2}{\varphi_1^2}CR_{t-1}^1 + \frac{\varphi_1^1\varphi_2^2}{\varphi_1^2}CC_{t-1}^1 + [\frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{1t}^1 - \epsilon_{1t}^2) + \epsilon_{2t}^2]. \tag{IV.48}$$

First note that $CR_{t-1}^1$ and $CR_{t-2}^2$ cannot be the same process since in that case we would have only one cointegration relationship between three series, what cannot be detected testing in a pairwise fashion. Note also that the cointegration relationship between $X_{2t}$ and $X_{3t}$ — say, $CR_t^3$ — is a linear combination of $CR_t^1$ and $CR_t^2$. Then, let:

$$CR_t^3 = b_1 CR_t^1 + b_2 CR_t^2.$$

Define now $a_1 = \frac{\alpha_1^1\varphi_2^2}{\varphi_1^2}$, and $a_2 = \alpha_2^2 - \frac{\alpha_1^2\varphi_2^2}{\varphi_1^2}$. In order to conclude that $X_{2t}$ and $X_{3t}$ have the $WF$ we should be able to write $\Delta X_{3t}$ in eq. (IV.48) as a linear combination of $CR_{t-1}^3$ and $CC_{t-1}^1$ plus white noise. Thus, we need:

$$(b_1 - a_1)CR_{t-1}^1 + (b_2 - a_2)CR_{t-1}^2 = dCC_{t-1}^1 + v_t, \tag{IV.49}$$

where $v_t$ is white noise. The cointegration relationships $CR_t^1$ and $CR_t^2$ are linear combinations of $X_{1t}$, $X_{2t}$ and $X_{3t}$. Therefore, if $d = 0$, eq. (IV.49) implies that there exists a linear combination

of $X_{1t}$, $X_{2t}$ and $X_{3t}$ that is white noise. As showed by Vahid and Engle (1993) and highlighted in §IV.2.2, the orthogonal complements of matrices $C(1)$ and $C_i^*$ (for $i > 0$) of the cointegrated VAR's MA representation — if they exist — are orthogonal. Therefore, a linear combination of the levels that delivers white noise is not possible in general. The only possibility is when the VAR in levels has only one lag, but we would be in the SCCF case. Hence, we focus on the case that $d \neq 0$. From eq. (IV.49) we can write:

$$CC_{t-1}^1 = \Upsilon \begin{bmatrix} CR_{t-1}^1 \\ CR_{t-1}^2 \end{bmatrix} + v_t^*, \tag{IV.50}$$

with $\Upsilon = [(b_1 - a_1),\ (b_2 - a_2)]/d$ and $v_t^* = v_t/d$.

Note now that if the three variables we are dealing with share a single $WF$ we can also write:

$$\Delta X_t = c + \alpha_{(3\times 2)} \begin{bmatrix} CR_{t-1}^1 \\ CR_{t-1}^2 \end{bmatrix} + \gamma CC_{t-1}^1 + \epsilon_t. \tag{IV.51}$$

Plugging eq. (IV.50) into eq. (IV.51) we get:

$$\Delta X_t = c + \alpha_{(3\times 2)}^* \begin{bmatrix} CR_{t-1}^1 \\ CR_{t-1}^2 \end{bmatrix} + \epsilon_t^*, \tag{IV.52}$$

with $\alpha^* = \alpha + \gamma\Upsilon$; and $\epsilon_t^* = \epsilon_t + \gamma v_t^*$. Since $\alpha^*$ has dimension $3 \times 2$, eq. (IV.52) implies a $SCCF$ structure. Hence, when analyzed in a pairwise fashion, *pure* $WF$ structures are not transitive.

A trivial exception to this statement would be when $b_1 = a_1$ *and* $b_2 = a_2$, what will imply that $CC_{t-1}^1$ is white noise (see eq. (IV.49)), so the first differences of the series are white noise except for the cointegration relationships.

In summary, the only definition of *'common cycles'* that is transitive when considered in a pairwise fashion is the original *SCCF*.

### IV.4.2 A strategy for dealing with WF (and PSCCF)

As noted in §IV.2.2, the number — $s$ — of *SCCF* cannot be larger than the number of common trends (i.e; $s \leq n - r$). Since in any subset of series inside a *fully cointegrated set* there is a single common trend ($n - r = 1$), a single *'common cycle'* ($n - s = 1$) can be possible between two of those series, but not more than two (because the conditions $s \leq n - r$, and $n - r = 1$

imply $n - s \geq n - 1$, so that $n - s$ can be 1 only if $n = 2$). To avoid this limitation one could consider *WF* structures but, as showed above, they are not transitive.

The lack of transitivity can be solved by using all the relevant cointegration relationships for the two series of a pair in a fully cointegratd subset. To see this, consider the following systems:

$$\begin{aligned} \Delta X_{1t} &= c_1^1 + \alpha_{11}^1 CR_{t-1}^1 + \alpha_{12}^1 CR_{t-1}^2 + \varphi_1^1 CC_{t-1}^1 + \epsilon_{1,t}^1, \\ \Delta X_{2t} &= c_2^1 + \alpha_{21}^1 CR_{t-1}^1 + \alpha_{22}^1 CR_{t-1}^2 + \varphi_2^1 CC_{t-1}^1 + \epsilon_{2,t}^1, \end{aligned} \tag{IV.53}$$

$$\begin{aligned} \Delta X_{1t} &= c_1^2 + \alpha_{11}^2 CR_{t-1}^1 + \alpha_{12}^2 CR_{t-1}^2 + \varphi_1^2 CC_{t-1}^2 + \epsilon_{1,t}^2, \\ \Delta X_{3t} &= c_2^2 + \alpha_{21}^2 CR_{t-1}^1 + \alpha_{22}^2 CR_{t-1}^2 + \varphi_2^2 CC_{t-1}^2 + \epsilon_{2,t}^2. \end{aligned} \tag{IV.54}$$

Equalizing the first lines of eq. (IV.53), and eq. (IV.54), solving for $CC_{t-1}^2$, and plugging the result in the second line of eq. (IV.54) we get:

$$\Delta X_{3t} = c^3 + \alpha_1^* CR_{t-1}^1 + \alpha_2^* CR_{t-1}^2 + \varphi^* CC_{t-1}^1 + \epsilon_t^*,$$

where $c^3 = \frac{c_1^1 - c_1^2}{\varphi_1^2}$; $\alpha_1^* = \alpha_{21}^2 + \frac{\varphi_2^2}{\varphi_1^2}(\alpha_{11}^1 - \alpha_{11}^2)$; $\alpha_2^* = \alpha_{22}^2 + \frac{\varphi_2^2}{\varphi_1^2}(\alpha_{12}^1 - \alpha_{12}^2)$; $\varphi^* = \frac{\varphi_2^2}{\varphi_1^2}\varphi_1^1$; and $\epsilon_t^* = \epsilon_{2t}^2 + \frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{2t}^1 - \epsilon_{1t}^2)$. Thus, after including all relevant cointegration relationships, *WF* becomes transitive. In general, for any $n > 3$, assuming that all series share a unique common cycle as defined by *WF*, the full model would be:

$$\Delta X_t = c + \alpha\beta' X_{t-1} + \delta_\perp CC_{t-1} + \epsilon_t, \tag{IV.55}$$

where $X_t$ is a $n \times 1$ vector; $\alpha$ and $\beta$ are $n \times r$; and $\delta_\perp$ is $n \times 1$. For deriving the model for any pair $(X_{it}, X_{jt})$ define the $2 \times n$ 'selection matrix', $S = [(0, 0, ..., 1, 0, ...0)', (0, 0, ..., 1, 0, ...0)']'$ — which has ones only in positions $(1, i)$ and $(2, j)$ — and multiply both sides of eq. (IV.55) by $S$, to get:

$$\begin{bmatrix} \Delta X_{it} \\ \Delta X_{jt} \end{bmatrix} = \begin{bmatrix} c_i \\ c_j \end{bmatrix} + \begin{bmatrix} \alpha_{i1} CR_{t-1}^1 + ... + \alpha_{ir} CR_{t-1}^r \\ \alpha_{j1} CR_{t-1}^1 + ... + \alpha_{jr} CR_{t-1}^r \end{bmatrix} + \begin{bmatrix} \delta_{\perp,i} CC_{t-1} \\ \delta_{\perp,j} CC_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{i,t} \\ \epsilon_{j,t} \end{bmatrix}. \tag{IV.56}$$

Although eq. (IV.56) cannot be estimated, if the cointegration relationships are pairwise detectable, we could proceed in two steps: i) Apply the pairwise procedure to discover and

estimate all the cointegration relationships. ii) Perform *WF* tests between all the $N(N-1)/2$ pairs of series, including in each test all the estimated cointegration relationships that are relevant for each series in the pair. In order to avoid repetitions, for pairs of series inside a *fully cointegrated subset*, at most $n_1 - 1$ of the pairwise cointegration relationships inside the subset should be considered.

In §IV.5.2.2 we study this two step strategy by Monte Carlo.

**Remark 4** *This two steps strategy is also valid for $PSSCF(1)$, with the only modification that we should include the first lag of all relevant variables instead of the cointegration relationships.*

However, for WF structures, this two steps strategy raises three issues: a) We are not considering cointegration relationships which are not detectable with the pairwise strategy. b) What nominal size should be used for the cointegration tests between series outside the *fully cointegrated* subsets? c) For series inside some *fully cointegrated* subset, not all the $n_1 - 1$ cointegration relationships are necessarily relevant for both series, then, we could be introducing noise in the *WF* tests if cointegration relationships are not carefully selected.

Regarding the first issue, let $\alpha$ to be the matrix containing the equilibrium adjustment coefficients and $\Psi'_W$ the matrix that generates the cycles (see, eq. (IV.27)). If the *WF* restrictions are *purely weak*, in the sense that they are not also *strong*, not including relevant cointegration relationships will lead to over-reject the null of $s > 0$, in favor of $s = 0$. This because the left null spaces of $\alpha$ and $\Psi'_W$ do not intersect, and a matrix $\delta$ such that $\delta'\Psi'_W = 0$ *and* $\delta'\alpha = 0$, does not exist. Hence, we could lose some relevant restrictions. In a Monte Carlo experiment, Hecq et al. (2006) analyze the consequences of under-specifying the cointegration rank for WF tests. As one of the possibilities, they work with a three dimensional model with $s = r = 2$, and show that fixing $r$ at 1 leads to conclude $\hat{s} = 0$ in 5.8% of the replications and $\hat{s} = 1$ in the remaining 95% (see Table 4 in Hecq et al. (2006)).

This issue is mitigated by the relaxation procedure described in §IV.4.3.2, by which series that failed to show a common cycle with *all* the other series in the subset are allowed to enter it.

Consider now the second problem. In §II.2.1 we studied the problem of multiple testing when the objective was to detect cointegration relationships inside an *unknown* fully cointegrated subset, and concluded that regular nominal sizes could be used. This conclusion is not true when the focus is on all cointegration relationships — not only those within a fully cointegrated

subset. The main reason is that we do not have the *safety net* of requiring cointegration with (almost) all the series in the subset in order to accept the test's result as 'good' ones. If the multiple testing issue is to be mitigated, tighter significance levels should be used, at the cost of power deterioration.

Hence, there is a trade-off between correcting the *WF* tests with *wrong* cointegration relationships, and missing some relevant relationships. Again, if the *WF* restrictions are *purely weak*, not including relevant cointegration relationships will lead to over-reject the null of $s > 0$ in favor of $s = 0$. On the other hand, Hecq et al. (2006) show that over-specifying the cointegration rank is almost costless when it is 1, but is — wrongly — fixed at 2. Thus, tightening cointegration tests' nominal sizes for pairs outside *fully cointegrated subsets* does not seem, in principle, to be a critical requirement.

There is, however, an important difference with respect to Hecq et al. (2006). While in their case there is only one *wrong* cointegration relationship, in ours, the number could be large. This difference may be relevant in short samples. We delay the study of this issue to §IV.5.2.2.

Focus now on the third problem. As Hecq et al. (2006) argue, their result about over-specifying the cointegration rank comes from the fact that the coefficient of a non significant $I(1)$ variable in a $I(0)$ model converges in probability to zero. Note, however, that this argument does not apply to problem c), because the non-relevant cointegration relationships could be $I(0)$ in this case. The consequences of this issue are therefore unknown, and we delay its study to §IV.5.2.2.

### IV.4.3 Asymptotic properties

Our general framework can be summarized in the following assumptions:

**Assumption 5** *Assumptions 1 to 4 of chapter II hold.*

**Assumption 6** *There is, at least, one subset of sc components that share a single (possible weak) common cycle.*

**Assumption 7** *After correcting for possible cointegration relationships, $\Delta X_{it}$ is not white noise for $i = 1, ..., N$.*

Assumption 6 gives relevance to our objective of discovering single-cycle subsets, and assumption 7 rules out the 'trivial' common cycles that will appear if some components are white noise.

**Remark 5** *Instead of assumption 7 we could require that at least sc components satisfy that assumption, and the remaining ones may or may not satisfy it. This flexibilization would require testing the significance of the estimated common cycles' coefficients ($\delta_\perp$). This is how we proceed in the empirical applications of chapter VI.*

As in the case of cointegration tests, the pairwise procedure for discovering blocks of series sharing single common cycles may be subject to the problem of multiple testing. That is, since we are performing a large amount of tests, we may be inflating the false rejection probability (see §II.2.1.1 for a more detailed description of the multiple testing problem).

Let $SC$ denote a 'Single Cycle subset', i.e., a subset in which all series share a single cycle as defined by $WF$ (we will use $SC$ both, as the name of the subset, and to indicate the number of series inside it). As we did in chapter II, we study the multiple testing problem for three different types of pairs: (i) Pairs between two series inside the same $SC$ subset. Calling $s^*$ the true number of common cycles restrictions, for these pairs $s^*$ is 1. (ii) Pairs between a series inside and a series outside a SC subset ($s^* = 0$). (iii) Pairs between two outsiders ($s^* = 0$).

For the pairs in (i) the true hypothesis is $s = 1$, so that the multiple testing problem may lead to over reject $s > 0$ in favor of $s = 0$. For the other two types of pairs, under assumption 7, the true hypothesis is $s = 0$, so that, asymptotically, there cannot be a multiple testing issue, for asymptotic power is equal to one.

Therefore, in contrast with cointegration tests, we can focus only in pairs inside $SC$. The key difference is in the null hypothesis; while in the Johansen's test the first null is *no cointegration*, in common cyclical features, the first null is *at least one common cycle restriction.* Therefore, asymptotic power equal to one (the asymptotic probability of concluding $s = 0$ when it is true) ensures that — in large samples — we will not include wrong elements in the estimated $SC$.

#### IV.4.3.1 False rejection of $s > 0$

If the tests were independent, the probability of finding one common cycle between all series in $SC$ would be $(1 - \varphi)^{SC(SC-1)/2}$, where $\varphi$ is the nominal size of the individual tests. This probability quickly decreases with $SC$. If the tests are not independent this is not true, and the probability of discovering $SC$ depends on the degree of the dependence.

For Johansen's tests we showed that, far from being independent, tests inside a 'fully cointegrated' subset are asymptotically equivalent in the sense that the probability of finding the same result in all the tests is close to one in large samples ( theorem 1 of chapter II). The key for this result is that the unit root process that dominates the tests statistics' behavior is the same for all the pairs. Note that the only requirement for this result to be valid is that all series share the same trend. It does not make use of the dependence derived from the fact that we are testing all the possible pairs within a set of series (e.g., the test between $X_{1t}$ and $X_{2t}$ is not independent from the tests between $X_{1t}$ and $X_{it}$, or the tests between $X_{2t}$ and $X_{jt}$).

In cyclical features tests there are not unit root processes and we do not have something similar to theorem 1. As we argue below, test's dependence is determined by the 'similarity' between the pairs. We study this issue by means of a small Monte Carlo experiment. We want to distinguish two possible sources of dependence: first what we call *dependence a*, which is that derived from the existence of a unique common cycle; second, *dependence b*, is the associated with the fact that we are considering all the possible pairs in a set of series (i.e., using the same series in more than one test).

***Dependence a***

In order to keep things simple, at this stage, we deal only with $SCCF$ in stationary variables. For the case of $WF$, as the only difference is the inclusion if the cointegration relationships found in a previous step, the basic conclusions will not change. Anyway, the case of $WF$ adds additional complications, and we study them by Monte Carlo in §IV.5.2.

The DGP is similar to that in Candelon et al. (2005):

$$X_t = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & -0.4 \end{bmatrix} \Delta X_{t-1} + \epsilon_t, \tag{IV.57}$$

where $c_i$ is taken from the uniform distribution with limits 5 and 10; and $\epsilon_t \sim N(0, I_2)$. The coefficients' matrix of the corresponding to the first lag implied by eq. (IV.57) is:

$$\begin{bmatrix} 0.56 & -0.32 \\ -0.42 & 0.24 \end{bmatrix}.$$

In each of the 2000 Monte Carlo replications we generate $CC_{t-1} = [0.7,\ -0.4]\Delta X_{t-1}$ and

construct $Q$ vectors $X_t^*$ as:

$$X_t^* = \begin{bmatrix} c_1^* \\ c_2^* \end{bmatrix} + CC_{t-1} + \epsilon_t^*, \tag{IV.58}$$

where $c_i^*$ is taken from the uniform distribution with limits 5 and 10; $\epsilon_t^* = \sqrt{\theta}\ \epsilon_t + \sqrt{1-\theta}\ \eta_t$; and $\eta_t \sim N(0, I_2)$, independent from $\epsilon_t$. Hence, $Cov(\epsilon_t^*) = \theta(\epsilon_t \epsilon_t') + (1-\theta)(\eta_t \eta_t') = I_2$.

The parameter $\theta$ controls the 'similarity' between the $Q$ pairs within each replication. Indeed, $Cov(\epsilon_{it}^*, \epsilon_{jt}^*) = E[(\sqrt{\theta}\ \epsilon_t + \sqrt{1-\theta}\ \eta_{it})(\sqrt{\theta}\ \epsilon_t + \sqrt{1-\theta}\ \eta_{jt})'] = \theta I_2$, for $i \neq j$, where the indexes $i, j$ refer to different vectors of the same replication. The largest $\theta$ is, the more 'similar' the $Q$ vectors of the replication are. In the extreme case of $\theta = 1$, all the vectors are the same except for the constant.

Table IV.1 compares the cumulative distribution functions of the maximum of $Q$ independent $\chi_1^2$ — the distribution of each of the bivariate tests — random variables, and the maximum test statistic of the $Q$ bivariate systems of eq. (IV.58). We focus on the maximum because, for concluding that all the series share a single common cycle, we need to keep the null of $s > 0$ in all the pairs, what will happen if and only if the maximum test statistic is smaller than the critical value. Block (a) of the table gives the probability of not rejecting the $Q$ null hypothesis if the tests statistics were independent, and regular critical values used. Likewise, block (b) contains the probability of not rejecting all the $Q$ null hypothesis in our Monte Carlo experiments when regular critical values are used.

The comparison between blocks (a) and (b) shows that having the same single cycle is not a sufficient condition for the $SCCF$ tests statistics of $Q$ bivariate systems to be dependent. Indeed, when $\theta = 0$, the tests are independent (compare cumulative probabilities in block (a) with those of block (b) under $\theta = 0$). The degree of dependence increases with the 'similarity' between the bivariate systems, as the higher figures under $\theta = 0.7$ and $\theta = 0.95$ indicate.

***Dependence b***

We now analyze how previous results may change when the tests are performed for all the pairs of a set of series. For each pair $(X_t^*)$, we store the first series (say, $X_{1t}^*$), and we do this for the first $n$ — out of the total $Q$ — pairs of each replication. Thus, for each replication, we have a set of $n$ series, and we perform $SCCF$ tests between all the $n(n-1)/2$ possible pairs. Given the way we constructed the series, the covariance matrix of the residuals of the bivariate models is given by:

Table IV.1: Dependence between *SCCF* tests in bi-variate systems with the same 'common cycle'. Cumulative probabilities at point x.

| **x** | **0.45** | **1.64** | **2.71** | **3.84** | **6.63** |
|---|---|---|---|---|---|
| $\chi^2_1$ | **0.50** | **0.80** | **0.90** | **0.95** | **0.99** |
| **(a) Maximum from $Q$ independent $\chi^2_1$** | | | | | |
| $Q = 10$ | 0.00 | 0.11 | 0.35 | 0.60 | 0.90 |
| $Q = 45$ | 0.00 | 0.00 | 0.01 | 0.10 | 0.64 |
| $Q = 1225$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **(b) Maximum statistic of the $Q$ simulated $X^*_t$** | | | | | |
| **$\theta = 0$** | | | | | |
| $Q = 10$ | 0.00 | 0.09 | 0.31 | 0.56 | 0.90 |
| $Q = 45$ | 0.00 | 0.00 | 0.01 | 0.10 | 0.64 |
| $Q = 1225$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **$\theta = 0.7$** | | | | | |
| $Q = 10$ | 0.01 | 0.25 | 0.49 | 0.73 | 0.92 |
| $Q = 45$ | 0.00 | 0.02 | 0.20 | 0.44 | 0.80 |
| $Q = 1225$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 |
| **$\theta = 0.95$** | | | | | |
| $Q = 10$ | 0.19 | 0.63 | 0.82 | 0.89 | 0.97 |
| $Q = 45$ | 0.06 | 0.49 | 0.74 | 0.86 | 0.95 |
| $Q = 1225$ | 0.00 | 0.25 | 0.54 | 0.73 | 0.91 |

Block (a) includes probability of not rejecting the $Q$ null hypothesis if the tests statistics were independent, and regular critical values used.
Block (b) includes the probability of not rejecting all the $Q$ null hypothesis in our Monte Carlo experiments, for different values of the 'similarity parameter' ($\theta$, see eq. (IV.58)), when regular critical values are used.

$$\Omega_i = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}, \quad for \;\; i = 1, ..., n.$$

In order to preserve the residuals covariance matrix equal to the identity, we perform the experiments only for $\theta = 0$. As table IV.2 shows, figures are now somehow larger than under independence, but still, cumulative probabilities are far from the single test case. This result confirms that the degree of dependence between *SCCF* tests may be low even when they are performed between all the pairs in a set of variables.

#### IV.4.3.2 A discussion on test's dependence: relaxation to *almost full connection*

As aforementioned, the probability of finding a single 'common cycle' restriction between *every* possible pair in a set of *SC* series may go to zero very fast as *SC* increases. This is is an undesirable property for our approach. Note, however, that finding a *single cycle set* containing

Table IV.2: Dependence between *SCCF* tests between all possible pairs in a set of $n$ series with a single 'common cycle'. Cumulative probabilities at point x.

| **x** | **0.45** | **1.64** | **2.71** | **3.84** | **6.63** |
|---|---|---|---|---|---|
| $\chi^2_1$ | **0.50** | **0.80** | **0.90** | **0.95** | **0.99** |
| **(a) Maximum from $Q$ independent $\chi^2_1$** | | | | | |
| $n = 5$, *(10 pairs)* | 0.00 | 0.11 | 0.35 | 0.60 | 0.90 |
| $n = 10$, *(45 pairs)* | 0.00 | 0.00 | 0.01 | 0.10 | 0.64 |
| $n = 50$, *(1225 pairs)* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **(b) Maximum statistic of the $Q$ simulated $X^*_t$ $\theta = 0$** | | | | | |
| $n = 5$, *(10 pairs)* | 0.00 | 0.15 | 0.40 | 0.63 | 0.91 |
| $n = 10$, *(45 pairs)* | 0.00 | 0.00 | 0.03 | 0.16 | 0.70 |
| $n = 50$, *(1225 pairs)* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Block (a) includes probability of not rejecting the $Q$ null hypothesis if the tests statistics were independent, and regular critical values used.
Block (b) includes the probability of not rejecting all the $Q$ null hypothesis in our Monte Carlo experiments when regular critical values are used (only for $\theta = 0$, see eq. (IV.58)).

a large proportion of the *SC* series is also of interest. In such a case, we will correctly impose the common cycle restriction to series that truly have it, but we will not impose the restriction to some other series that also have it. This will not bias the forecasts, but may make them less efficient.

Thus, as we did for *full cointegration*, we can relax the single-cycle requirement for *all* the pairs, and require that *almost* all the tests lead to the conclusion $s = 1$. There are, however, two relevant differences with respect to the relaxation to *almost full cointegration*. First, while in that case the relaxation was a small samples correction, now, it may be necessary even in large samples because tests are not equivalent. Second, the relaxation strategy is asymptotically costless, for the probability of finding a common cycle when there is none, is asymptotically zero (asymptotic power is 1)[4].

Our approach to study the properties of the relaxation procedure applied to common cycles is as follows. When performing the $C^n_2 = n(n-1)/2$ tests, the individual asymptotic probability of not rejecting the null of $s > 0$ for each of the pairs formed by two series of *SC*, is $1 - \varphi$. For any other pair, this probability is zero.

[4]This means that in infinitely large samples (and fixed *SC*) we could use a significance level as small as we need in order to fix the probability of discovering *SC*, even under independence. For example, for $SC = 15$ there are 105 pairs, so if we want the probability of discovering *SC* to be, say, 0.95 we should set the significance level of the individual tests at $\varphi = 0.95^{1/105} - 1 = 0.0005$. This value is too small to be used in practice, even for relatively long time series. This is why, although strictly speaking the relaxation procedure is not needed asymptotically, we still consider it necessary even for large samples.

In terms of the random graph theory, after performing the $C_2^n$ tests, we have a graph with $n$ vertices and $C_2^n$ possible edges (connections). Within the $C_2^n$ possible edges, $C_2^{sc} = SC(SC-1)/2$ of them have probability $(1-\varphi)$ and the remaining $C_2^n - C_2^{sc}$ have probability zero. Finding the largest *single-cycle subset*, is equivalent to find the largest *fully connected* subgraph — i.e., the largest subgraph in which all possible edges are present. This is the maximum *clique problem* described in the random graph literature (see, *interalia*, Matula (1976); Derényi et al. (2005); and Newman (2009)).

Since edges probabilities for pairs in which at least one series is not in $SC$ are zero, we can focus on the subgraph formed by the series in $SC$. Under independence, this is a standard Erdós-Rényi random graph with edge probability $p = 1 - \varphi$. Letting $\hat{SC}$ to be the size of the maximum clique, we are interested in computing the probability function, $P(\hat{SC} = k)$, $\forall$ $0 \leq k \leq c$.

It tuns out that finding a general solution for $P(\hat{SC} = k)$ — for any $2 \leq k \leq SC$ — is quite a hard problem. In fact, to the best of our knowledge, this problem is not solved in the random graph literature. Hence, we give a *second best* description of the pairwise procedure applied to common cycles based on available measures related to $P(\hat{SC} = k)$, and on Monte Carlo evidence.

Derényi et al. (2005) give the *threshold probability for the k-clique percolation* — $p_c(k)$. The authors define two cliques of size $k$ to be *adjacent* if they share $k-1$ vertex, and a *k-clique chain* as the union of a sequence of adjacent *k-cliques*. Next, they define two k-cliques to be *connected* when they are part of the same *chain*. Using these definitions, the *k-clique percolation cluster* is defined as the maximal *k-clique connected* subgraph. Hence, the *k-clique percolation cluster* is the union of all k-cliques which are *connected* to a particular k-clique. The *k-clique percolation probability*, $p_c(k)$, is defined as the threshold probability such that, for $p \geq p_c(k)$ the random graph will converge almost sure to a *k-clique cluster*, and for $p < p_c(k)$ the probability of having *k-clique cluster* is almost zero. The authors show that:

$$p_c(k) = \frac{1}{[(k-1)sc]^{\frac{1}{k-1}}}, \tag{IV.59}$$

where $sc$ is the number of vertices, see Derényi et al. (2005) and Palla et al. (2007) for the details.

Using eq. (IV.59), we can fix a proportion $\delta$ and chose $k = \delta(SC-1)+1$, so that $p_c(k)$

will give the edge probability required to find a subset of size $SC$, in which, for each series, a common cycle was found with, at least, $\delta(SC-1)$, out of the $SC-1$ remaining series. In other words, if the probability of not rejecting the — true — null is larger than or equal to $p_c(\delta(SC-1)+1)$, we — almost surely — will find a subset of size $SC$ in which, for each series, a single common cycle was found with at least $\delta(SC-1)$ of the other series.

Figure IV.1 shows this threshold probability for different choices of $\delta$ and $SC$. These probabilities can be seen as the required magnitude for $(1-\varphi)$ for finding the *almost fully connected* graph we are looking for.

For instance, for $SC = 40$ and $\delta = 0.8$, we would need $(1-\varphi) = 0.8$. This implies that relaxing the requirement from *full connection*, to *almost full connection* may lead to a great improvement in the probability of finding the true $SC$.

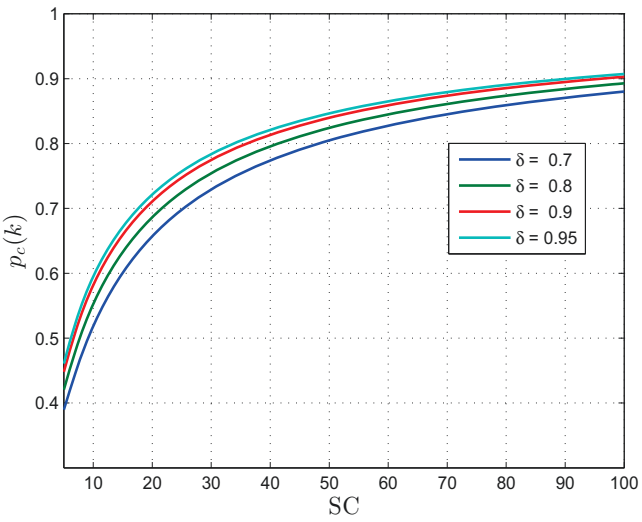



Figure IV.1: Percolation probabilities $p_c(k)$, for $k = \varphi(SC-1)+1$

Note, however, that eq. (IV.59) is valid asymptotically — large $(SC-k)$ —, so it may be a quite rough measure in our cases of interest, as we are interested in lagre values of $\delta$. To better understand the properties of our procedure, we perform a small simulation study. We simulate a random graph of size $N$ with independent edge probability $p$. Then, we apply the algorithm to find the largest *almost fully connected* subset described in chapter III (see, §III.5). Results are included in table IV.3. As a general conclusion: the relaxation procedure, which is asymptotically costless, allows to reach acceptable results even when regular significance levels are used. For example, with a nominal significance level $\varphi = 0.05$, when the true size of the set is 50, the expected ratio *size of the estimated subset* to *true size of the subset* is 0.99. Recall

that, as we argued §IV.4.3.1, tests' independence is an extreme assumption.

Table IV.3: Formation of *almost fully connected* subsets

| | | **N=5** | **N=10** | **N=20** | **N=25** | **N=40** | **N=50** |
|---|---|---|---|---|---|---|---|
| | | | | $\boldsymbol{p = 0.95}$ | | | |
| *Full contection* | mean($\hat{N}/N$) | 0.92 | 0.83 | 0.73 | 0.69 | 0.61 | 0.56 |
| | mean($I_{\hat{N}=N}$) | 0.62 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Almost Full conection* | mean($\hat{N}/N$) | 0.99 | 0.95 | 0.96 | 0.98 | 0.97 | 0.99 |
| | mean($I_{\hat{N}=N}$) | 0.94 | 0.56 | 0.40 | 0.56 | 0.28 | 0.49 |
| | | | | $\boldsymbol{p = 0.99}$ | | | |
| *Full contection* | mean($\hat{N}/N$) | 0.98 | 0.96 | 0.92 | 0.90 | 0.86 | 0.83 |
| | mean($I_{\hat{N}=N}$) | 0.90 | 0.61 | 0.16 | 0.04 | 0.00 | 0.00 |
| *Almost Full conection* | mean($\hat{N}/N$) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | mean($I_{\hat{N}=N}$) | 1.00 | 0.97 | 0.98 | 1.00 | 1.00 | 1.00 |

- The relaxation parameter is 1 for $N = 5$, $N = 10$; 2 for $N = 20$; and 5 for $N = 50$.
- 'mean' denotes the mean across experiments.
- The number of experiments is 1000.
- $\hat{N}$ is the number of series included in the largest (almost) fully connected subset.
- $I_{\hat{N}=N}$ denotes the indicator function that takes the value 1 if $\hat{N} = N$ and 0 otherwise.

### IV.4.4 On the bivariate sub-models

The pairwise strategy applied to common cycles consists of testing for a common cycle in all possible pairs of series and look for the largest subset in which *almost all* the pairs have the cycle. This strategy requires estimating partial models and it could be thought to imply a power loss with respect to a 'complete' model approach (when feasible). To analyze this issue we perform a small simulation study.

We consider two alternative DGPs. Both of them are N-dimensional stationary VAR models in which a subset of $SC$ series share a single SCCF. The DGPs differ in how the common cycle is generated.

#### IV.4.4.1 Stationary DGPs

Both DGPs have the same general structure:

$$X_t = c + \Pi X_{t-1} + \epsilon_t, \tag{IV.60}$$

where $\epsilon_t \sim N(0, I)$; and the roots of $det(I - \Pi L)$ are all outside the unit circle. We want to simulate a situation in which only a subset $SC$ of series share a single cycle and there are no

more common cycles restrictions in the system. In order to simplify the system's dynamics, we assume that the series not belonging to $SC$ have zero covariance with all the other series. Thus, matrix $\Pi$ has the following structure:

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1,sc} & 0 & 0 & \dots & 0 \\ \vdots & & & & & & & \\ \pi_{sc,1} & \pi_{sc,2} & \dots & \pi_{sc,sc} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \pi_{sc+1,sc+1} & 0 & \dots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \pi_{N,N} \end{bmatrix},$$

where the upper left $SC \times SC$ sub-matrix is equal to $\delta_{\perp}^{*}\Psi^{*'}$; and $\delta_{\perp}^{*}$ and $\Psi^{*}$ are $SC \times 1$ vectors. Partition the vector $X_t$ in its first $SC$ elements and the remaining $N - SC$, and call $X_t^{sc}$ to the first sub-vector. Then, the common cycle is $\Psi^{*'}X_{t-1}^{sc}$, and $\delta_{\perp}^{*}$ contains the coefficients of the common cycle in each of the first $SC$ series.

Since we want eq. (IV.60) to be stationary, we need the roots of the characteristic polynomial $det(I - \Pi L)$ to be outside the uni circle, or equivalently, the eigenvalues of the companion matrix inside it. As the model has only one lag, the companion matrix is $\Pi$ and its eigenvalues are the solution of $det(\Pi - \lambda I) = 0$. To solve this problem we write:

$$\Pi - \lambda I = \begin{bmatrix} A_{sc\times sc} - \lambda I & B_{sc\times(n-sc)} \\ C_{(n-sc)\times sc} & D_{(n-sc)\times(n-sc)} - \lambda I \end{bmatrix},$$

where $B$ and $C$ are full of zeros; $D$ is diagonal; and:

$$A = \begin{bmatrix} d_1\psi_1 & d_1\psi_2 & \dots & d_1\psi_{sc} \\ \vdots & & & \\ d_{sc}\psi_1 & d_{sc}\psi_2 & \dots & d_{sc}\psi_{sc} \end{bmatrix},$$

where $d_i$ and $\psi_i$ are the elements of $\delta_{\perp}^{*}$ and $\Psi^{*'}$, respectively. Using the result in Silvester (2000) about determinants of block matrices:

$$det(\Pi - \lambda I) = det[(A - \lambda I) - B(D - \lambda I)^{-1}C]\ det[D - \lambda I].$$

Since $B$ and $C$ are full of zeros we get that;

$$det(\Pi - \lambda I) = det[A - \lambda I]\ det[D - \lambda I].$$

Therefore, the solutions of $det(\Pi - \lambda I) = 0$ are the solutions of $det(A - \lambda I) = 0$, *and* the solutions $det(D - \lambda I) = 0$, which are the eigenvalues of $A$ and $D$, respectively. In other words, the roots of the characteristic polynomial $det(I - \Pi L)$ are the inverse eigenvalues of $A$ and $D$.

Since $D$ is diagonal, its eigenvalues are the elements of the diagonal, so that we need to set all the $\pi_{ii}$ smaller than 1 in absolute value, for all $i \geq sc + 1$.

Note now that sub-matrix $A$ has rank 1 so that it has only one non-zero eigenvalue. Thus, as the sum of the eigenvalues equals the matrix's trace, the non-zero eigenvalue is $\sum_{i=1}^{sc} d_i\psi_i$. Therefore, after setting $|\pi_{ii}| < 1$, $\forall i \geq sc + 1$, the stationarity condition of eq. (IV.60) is:

$$|\sum_{i=1}^{sc} d_i\psi_i| < 1.$$

There are infinitely many different possibilities of $\delta_\perp^*$ and $\Psi^*$ that would satisfy the stationarity condition. Three of them that may be of interest are:

**DGP 0** : $\delta_\perp^*$ is filled with uniform random values between 0.7, and 1 and $\Psi^*$ is filled with uniform random values between $\frac{1}{1.2SC}$ and $\frac{1}{1.1SC}$.

**DGP 1** : Same as DGP 0 but imposing $z$ zeroes in $\Psi^*$ so that we can change $SC$ in the denominator for $SC - z$ and the non zero entries will be larger. We set $SC - z = 2$, so that the common cycles will be generated by two of the $SC$ series.

**DGP 2** : Same as DGP 1 but allowing some negative entrances in $\delta_\perp^*$ (there will be some counter-cyclical variables). This allows to increase the non-zero entrances in $\Psi^*$ with respect to option DGP 1. We impose negative coefficients to 20% of the variables in $\delta_\perp^*$, hence, the number of non zero coefficients in $\Psi^*$ is $(2 + 0.2 \times SC)$.

Note that in DGP 0 all entries of vector $\psi$ will be rather small even for relatively small $SC$, so that quite large samples would be necessary in order to statistically distinguish them from zero. Thus, we focus only on DGP 1 and DGP 2.

#### IV.4.4.2 Monte Carlo results: pairwise vs full model approaches

For DGP 1 and DGP 2, we consider four different combinations of $[N, SC]$: $[6, 2]$; $[6, 3]$; $[9, 2]$; and $[9, 3]$. We therefore have eight different sub-DGPs. For each of these, we apply the canonical correlations test for SCCF in the full model and using the pairwise strategy. While in the first case we count the number of times that the test finds the correct number of common cycles, in the latter, we count the number of times that we find a common cycle in *all* the pairs that truly share the cycle. Additionally, for the pairwise approach, we count the number of times that we find a common cycle in all but one of the pairs that truly share the cycle. In both cases the nominal size of the tests is 1%.

For the full model the true lag length is known (one), and we use it. For the bi-variate models the lag length is unknown and we determine it using the AIC and the BIC. As it was the case in chapter II, the lag length depends on the *type of pair*. There are three different cases to distinguish: both series have the common cycle; only one of them have it; or, none of them have the common cycle. It can be shown that the lag length is non-increasing from the first to the third case (for a discussion on linear transformations of VAR processes see Lütkepohl (1984)).

The number of replications is 1000 and we consider three different sample sizes; $T = 100$, $T = 200$, and $T = 400$. Table IV.4 summarizes the results. For relatively large samples ($T = 400$ and $T = 200$), the full system approach delivers the expected results for all $N$ and $SC$ configurations. For the cases with $SC = 2$, the results of the pairwise approach are very close to those of the full sample procedure. The small differences are due to the fact that while in the latter case we are using the true lag length, in the former we need to estimate it. As expected, for larger $SC$ the pairwise approach deteriorates, but this deterioration is completely solved by admitting failure to find a common cycle in just one of the $SC \times (SC - 1)/2$ pairs.

For small sample sizes ($T = 100$), previous conclusions are not anymore valid. The probability of finding the true number of common cycles in the full model approach is close to 0.8 for $N = 6$, and it becomes lower than 0.3 for $N = 9$. In contrast, the pairwise approach only deteriorates for the case of $N = 9$ and $SC = 4$, for which the probability of finding a common cycle between *all* the pairs that truly have it (which are $4 \times 3/2 = 6$) decreases from around 0.9, for $T = 400$, to around 0.6, for $T = 100$. Again, this deterioration virtually disappears when we admit one of the pairs to fail in showing a common cycle.

As a conclusion, when common cycles are pairwise detectable, nothing is lost by proceeding

in a pairwise fashion. On the contrary, important power gains for finding the true number of common cycles in short samples can result from this procedure with respect to the full model approach.

Table IV.4: Probability of finding all common cycles' relationships. Comparison between the multivariate *Can Cor* test and the Pairwise procedure.

| | | | DGP | DGP 1 | | | | DGP 2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | N | 6 | | 9 | | 6 | | 9 | |
| | | | *SC* | 2 | 3 | 2 | 4 | 2 | 3 | 2 | 4 |
| $T$ =400 | PW | AIC | CanCorr | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 |
| | | | all | 0.98 | 0.94 | 0.97 | 0.89 | 0.98 | 0.93 | 0.98 | 0.91 |
| | | | all-1 | — | 1.00 | 1.00 | 0.99 | — | 0.99 | 1.00 | 0.99 |
| | | BIC | all | 0.99 | 0.97 | 0.98 | 0.93 | 0.99 | 0.96 | 1.00 | 0.94 |
| | | | all-1 | — | 1.00 | 1.00 | 1.00 | — | 1.00 | 1.00 | 1.00 |
| $T$ =200 | PW | AIC | CanCorr | 0.99 | 0.99 | 0.98 | 0.96 | 0.99 | 0.99 | 0.98 | 0.98 |
| | | | all | 0.98 | 0.94 | 0.97 | 0.85 | 0.97 | 0.94 | 0.98 | 0.84 |
| | | | all-1 | — | 1.00 | 1.00 | 0.98 | — | 1.00 | 1.00 | 0.98 |
| | | BIC | all | 0.99 | 0.96 | 0.98 | 0.89 | 0.99 | 0.97 | 0.99 | 0.88 |
| | | | all-1 | — | 1.00 | 1.00 | 0.99 | — | 1.00 | 1.00 | 1.00 |
| $T$ =100 | PW | AIC | CanCorr | 0.76 | 0.78 | 0.24 | 0.28 | 0.71 | 0.78 | 0.25 | 0.29 |
| | | | all | 0.99 | 0.92 | 0.97 | 0.64 | 0.97 | 0.92 | 0.97 | 0.62 |
| | | | all-1 | — | 1.00 | 1.00 | 0.95 | — | 0.99 | 1.00 | 0.94 |
| | | BIC | all | 0.99 | 0.95 | 0.98 | 0.66 | 0.99 | 0.95 | 0.99 | 0.64 |
| | | | all-1 | — | 1.00 | 1.00 | 0.97 | — | 1.00 | 1.00 | 0.95 |

- Number of replications: 1000.
- *CanCor* rows contain the probabilities of finding the correct number of common cycles relationships by means of the multivariate canonical correlations test, when all the $N$ variables are included in the model.
- *PW* rows contain the probabilities of finding a common cycle in all (and all but one) the pairs of series that truly share the cycle when the tests are done by the *CanCorr* test but in a pairwise fashion, and the lag length is selected according to the AIC (BIC) criteria.
- Nominal size of the tests: 0.01.

## IV.5 The behavior of the pairwise strategy for common cycles

We now turn to the analysis of the pairwise strategy in the framework for which it has been designed; relatively large N. In this section we perform some Monte Carlo experiments to fulfill three objectives: confirm the analysis of §IV.4.3, analyze the small sample properties of the pairwise strategy, and compare its performance with a DFM alternative. The same caveat of chapter II is valid here. Our goal is just to evaluate if the usual and simple strategy of

estimating factor models by principal components can be used in our framework of interest (large $N$ and relatively small single-feature subsets). Our objective is not making a general comparison between our approach and DFM. We do not want to extract general results.

As discussed in chapter II, the ideal procedure will: 1) from all the $N$ series, identify a large proportion of those that truly share the cycle (those in $SC$); 2) not include wrong series in the estimated subset(s), $\hat{SC}$. These conditions are closely related to what Castle et al. (2011) call *potency* and *gauge*, respectively. While gauge measures the retention frequency of irrelevant variables when selecting among a — potentially large — set of candidates, potency denotes the average retention frequency of relevant variables.

In chapter II we also argued that condition 2 is the most critical one, since including a large amount of wrong series would have disastrous consequences for the procedure. Hence, we analyze the two conditions giving special attention to the second one.

We split the analysis into two parts. In the first one (§IV.5.1), we focus on $N$-dimensional stationary process in which a subset of $SC$ series share a single common cycle (SCCF), and the other $N - SC$ variables have their own stationary dynamics.

In the second part (§IV.5.2), we generalize previous results to non stationary models with common cycles (WF) and common trends. Now, each of the series belongs to one of four groups: *only common trend*, *only common cycle*, *common trend and common cycle*, or *none*.

### IV.5.1 Stationary models

In this subsection we study the behavior of the pairwise approach applied to common cycles in a stationary framework. The analysis is composed by three parts; in §IV.5.1.1 we describe the DGPs, in §IV.5.1.2 and §IV.5.1.3 we study the behavior of the pairwise strategy, and §IV.5.1.4 we compare our procedure with a DFM alternative.

#### IV.5.1.1 Monte Carlo design for stationary models

We consider two alternative types of DGPs; two stationary VAR processes and a DFM.

The VAR processes are DGP 1 and DGP 2 analyzed in §IV.4.4.1. For the DFM (DGP 3) we consider the following model:

$$\begin{aligned} X_t &= \Lambda F_t + \epsilon_t, \\ F_t &= \rho F_{t-1} + \eta_t, \end{aligned} \tag{IV.61}$$

where $\Lambda$ is a $n \times 1$ vector of factor loadings whose first $SC$ elements are taken from the uniform distribution with parameters $[0.1, 0.8]$ and the others are 0; $\rho = 0.6$; $\epsilon_t \sim N(0, \Sigma)$; $\Sigma$ is diagonal with $\sigma_{ii}^2 = (1 - \frac{\lambda_i^2 \sigma_\eta^2}{1-\rho^2})$; and $\eta_t \sim N(0, I)$ independent of $\epsilon_t$. This structure implies that each series has unit variance, and the signal to noise ratio is $\lambda_i^2 \sigma_F^2 / (1 - \lambda_i^2 \sigma_F^2)$, with $\sigma_F^2 = \frac{\sigma_\eta^2}{1-\rho}$.

For the three DGPs we consider three scenarios and sample sizes. In scenario 1 we set $SC = 10$; in scenario 2, $SC = 25$; and in scenario 3, $SC = 40$. The sample sizes are $T = 100$, $T = 200$, and $T = 400$.

For each DGP, scenario and sample size, we perform 1000 Monte Carlo replications. In each replication we simulate a 100-dimensional model in which a subset of $SC$ series share a single common cycle. Our objective is to discover the series in $SC$. To do that, we perform SCCF tests on all the 4950 bi-variate VAR sub-models that exist among the 100 series. Thus, for a particular DGP, scenario and sample size, we have 4.95 million sub-models (4950 for each replication). Since we have three DGPs, three scenarios, and three sample sizes, we have $(3 \times 3 \times 3) \times 4.95 = 133.65$ million sub-models to estimate. Additionally, since the lag length for each of the 4950 sub-models of a particular replication is unknown, we select it with the AIC, admitting between one and five lags. The high complexity of the simulations led us not to consider other DGPs. However, this does not imply that our results are valid only for VAR(1)processes because the analysis of §IV.4.3 is valid for any lag length.

#### IV.5.1.2 Monte Carlo results for stationary models

Let $Z_1$ be the number of correct series included in $\hat{SC}$ (the estimated subset of variables that share a unique common cycle). In addition to computing the potency, we compute $P(Z_1 \geq x)$ for $x \in [0, 1, ..., SC]$. The larger this probability for each $x$ is, the better the procedure's performance. Given the analysis in §IV.4.3, we expect $P(Z_1 = SC)$ to be close to 0 for moderately large $SC$ (even in large samples). However, we also expect $P(Z_1 = \delta SC)$ to be large, for $\delta$ close to, but smaller than one.

Likewise, let $Z_2$ be the number of incorrect series included in the estimated single-cycle set. Apart from evaluating the gauge, we compute $P(Z_2 \geq x)$ for $x \in [1, ..., N - SC]$. The larger this probability is, the worse the procedure's performance. Given the analysis in §IV.4.3, we expect $P(Z_2 > 0)$ to be close to 0 in large samples.

Given that the residuals' covariance matrix is diagonal, the only source of dependence among the $N(N-1)/2$ tests comes from the fact that each series is used in more than one test — what

we called *dependence b* in §IV.4.3. As showed there, this source of dependence is weak, so that the results of the pairwise approach are expected to be somewhat better, but close to the case of independence.

Figure IV.2 includes the plots for $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for scenario 2 and DGP 1. As it shows, the pairwise procedure performs as expected. Both, the probability of including *all* the correct series, and the probability of including wrong series are close to zero. For $T = 400$ and nominal size of the test 1% ($\varphi = 0.01$), on average, we correctly assign to $\hat{SC}$ $22.7/25 = 91\%$ of the true series. This figure reduces to $18.9/25 = 75.6\%$ for $\varphi = 0.05$. As expected, these proportions are somewhat larger than the corresponding ones of table IV.3 — 90% and 69%, respectively — since in that table tests were truly independent, as *dependence b* was eliminated.

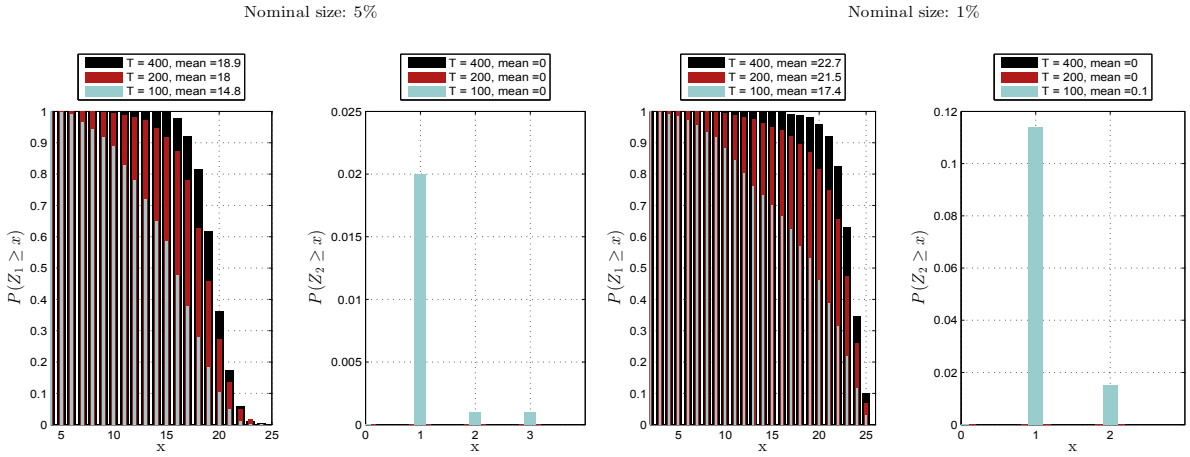



Figure IV.2: Pairwise approach. $P(Z_i \geq x)$. Scenario 2 ($SC = 25$). DGP 1.

The probability of including wrong series is very close to zero. This is so even in the case of $T = 100$ and $\varphi = 0.01$, in which, in almost 88% of the experiments no wrong series are included in $\hat{SC}$. For $\varphi = 0.05$, in 98% of the experiments no wrong series are included in $\hat{SC}$. For larger sample sizes we never include wrong series.

Table IV.5 summarizes the outcomes for all scenarios and sample sizes. Results are in line with the discussion of §IV.4.3: gauges are close to zero, and potencies deteriorate as $SC$ increases. As expected, reducing the nominal size from 5% to 1% is almost costless, except for $T = 100$ and scenario 3, where the gauge increases from 0 to 0.4%. This means that, on average, we include $0.004 \times 60 = 0.24$ wrong series in $\hat{SC}$. This still low gauge implies that in order to improve potencies, the strict *full connection* requirement can be relaxed as analyzed in §IV.4.3.2.

Table IV.5: Gauge and Potency of the Pairwise approach (in %). *Full connection, DGP 1.*

| | **Pairwise** (5%) | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 0.1 | 73.1 | 0.0 | 59.0 | 0.0 | 52.0 |
| T=200 | 0.0 | 84.4 | 0.0 | 72.0 | 0.0 | 65.8 |
| T=400 | 0.0 | 86.3 | 0.0 | 75.7 | 0.0 | 70.2 |

| | **Pairwise** (1%) | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 0.1 | 82.4 | 0.2 | 69.6 | 0.4 | 53.6 |
| T=200 | 0.0 | 92.8 | 0.0 | 86.0 | 0.0 | 75.4 |
| T=400 | 0.0 | 95.2 | 0.0 | 90.9 | 0.0 | 84.8 |

- Number of experiments: 500.
- For scenarios 1 to 3, $SC$ is 10, 25 and 40 respectively.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- $Z_2$ = number of wrong series included in $\hat{SC}$
- $Z_1$ = number of correct series included in $\hat{SC}$
- $Nexp$ = number of experiments

#### IV.5.1.3 Relaxation to 'almost full' connection

As argued in §IV.4.3, common cycles' tests inside a 'single-cycle' subset may be close to independence, so that the probability of correctly keeping the null of $s > 0$ in *all* the pairs that have the common cycle may rapidly go to zero as the number of series increases. In §IV.4.3.2 we also argued that the potency of the pairwise procedure could be notably increased by relaxing the *full connection* requirement to *almost full connection*. The relaxation consists of allowing to enter in $\hat{SC}$ those series for which a common cycle with at most $\lambda$ of the series initially estimated fully connected subset was not found, at the original significance level $\varphi$, but it was at $\varphi^*$ ($\varphi^* < \varphi$, see §III.5 for a detailed description of the relaxation process).

This strategy will lead to increase the potency of the procedure but may also increase the risk of including wrong series in $\hat{SC}$. This risk will be relevant only in small samples, for asymptotic power of common cycles' tests is equal to one.

Let $Z_2^\lambda$ be the number of series that do not belong to $SC$ but tests indicated a common cycle with all but $\lambda$ of the series in the original $\hat{SC}$. That is, $Z_2^\lambda$ represents the number of potential candidates to enter the *almost fully connected* subset that we do not want to include. $Z_2^\lambda$ is

expected to be a decreasing function of the original $\hat{SC}$.

We now repeat the experiments of §IV.5.1.2 but applying the relaxation procedure. In table IV.6 we analyze the performance of that procedure for $T = 200$, $\varphi = 0.05$, and $\varphi^* = 0.005$[5]. The table shows the mean number of potential candidates for each scenario and relaxation parameter up to $\lambda = 3$. While column (a) contains the mean number of series that have between 1 and $\lambda$ holes in the current *(almost) fully connected set*, column (b) includes the series of column (a) whose holes were filled after relaxing tests to the 0.5%. Columns (c) and (d) are analog to (a) and (b) but wrong candidates are excluded.

*'Ratios'* columns show not only that in all scenarios all potential candidates are correct series, but also that requiring p-value>0.005 does not play any role in filtering wrong candidates, for all series in column (a) are correct ones. Additionally, the cost of requiring p-value>0.005 — instead of imposing no requirements — in terms of reducing the amount of correct candidates to be included in $\hat{SC}$ is quite low, as the small difference between columns (b) and (a) indicate.

Figure IV.3 details the gains of the relaxation for all scenarios and sample sizes. It shows that the benefits are substantial in all situations; for scenarios 3 and 1, and $T = 100$, the correct number of series increases from $0.52 \times 40 = 20.8$ to $0.79 \times 40 = 31.6$, and from 7.3 to 8.8, respectively.

As aforementioned, these gains are costless for $T = 400$ and $T = 200$. As Figure IV.4 shows, for $T = 100$, the costs are still very low for scenarios 2 and 3 but somewhat more important for scenario 1. Anyway, when applying the relaxation in scenario 1 with $T = 100$ we reach a ratio $Z_2/\hat{SC} = 0.02/0.88 = 2.3\%$, meaning that we have a $\hat{SC}$ with only 2.3% of wrong series.

In sum, the Monte Carlo evidence confirms that the relaxation procedure virtually solves the consequences of the multiple testing problem, with almost zero cost.

#### IV.5.1.4 Comparison with DFM: stationary models

In this section we compare the performance of the pairwise strategy with a DFM alternative. For implementing the DFM, we apply the traditional Principal Components procedure to the whole data set and keep the number of factors suggested by the information criteria $IC_k$ and the three penalty functions detailed in Bai (2004). When each penalty function suggest a different number of factors we choose the minimum, otherwise we chose the mode. This procedure implies

[5]Recall that though there were important differences between nominal sizes 1% and 5%, after the relaxation these differences virtually disappear so we present results only for $\varphi = 5\%$.

Table IV.6: Statistics of the Relaxation process. Mean number of potential candidates ($T = 200$, $\varphi = 0.05$, and $\varphi^* = 0.005$).

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 1$**

| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.005 | (c) No Rest | (d) pval>0.005 | c/a | d/b |
| Sce 1 | 0.86 | 0.81 | 0.85 | 0.81 | 1.00 | 1.00 |
| Sce 2 | 2.40 | 2.28 | 2.40 | 2.28 | 1.00 | 1.00 |
| Sce 3 | 3.73 | 3.56 | 3.73 | 3.56 | 1.00 | 1.00 |

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 2$**

| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.005 | (c) No Rest | (d) pval>0.005 | c/a | d/b |
| Sce 1 | 1.28 | 1.16 | 1.26 | 1.16 | 0.99 | 1.00 |
| Sce 2 | 4.28 | 3.94 | 4.28 | 3.94 | 1.00 | 1.00 |
| Sce 3 | 7.04 | 6.53 | 7.04 | 6.53 | 1.00 | 1.00 |

**Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 3$**

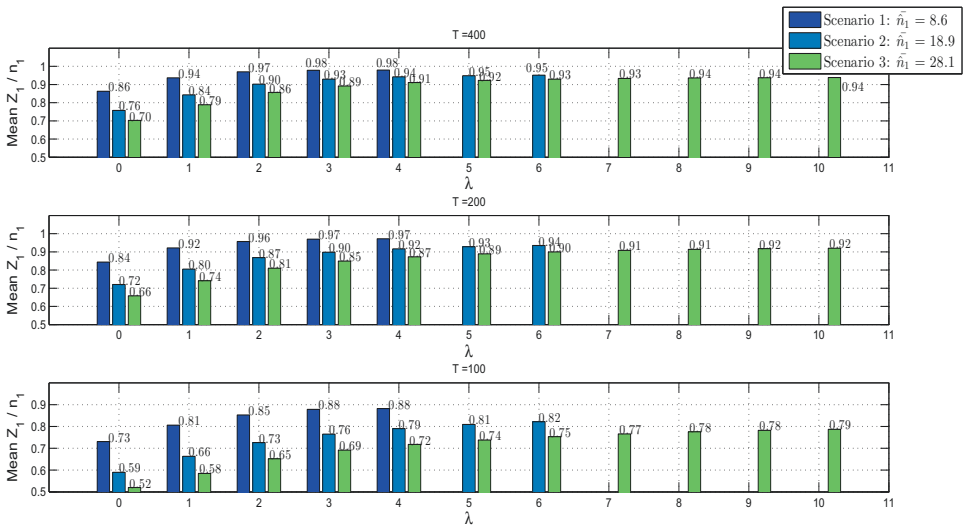
| | All Candidates ($Z^*$) | | Correct Candidates ($Z_1^*$) | | Ratios $Z_1^*/Z^*$ | |
|---|---|---|---|---|---|---|
| | (a) No Rest | (b) pval>0.005 | (c) No Rest | (d) pval>0.005 | c/a | d/b |
| Sce 1 | 1.52 | 1.27 | 1.40 | 1.27 | 0.99 | 1.00 |
| Sce 2 | 5.17 | 4.64 | 5.17 | 4.64 | 1.00 | 1.00 |
| Sce 3 | 8.83 | 8.07 | 8.83 | 8.07 | 1.00 | 1.00 |

- Number of experiments: 500. For scenarios 1 to 3, $SC$ is $10, 25$ and 40 respectively.
- The $\hat{SC}$ subset is updated in each step.
- Column (a) contains the mean number of series that have between 1 and $\lambda$ holes in the current *(almost) fully connected set* (common cycles restriction at 5% was rejected with at least $\lambda$ series in $\hat{SC}$).
- Series in column (b) and (d) are those of column (a) and (d) whose holes were 'filled' after relaxing common cycles tests to 0.5%.
- A comparison between columns (a) and (b) or columns (c) and (d) gives an idea of the effects of requiring common cycles restriction at 0.5% for the holes to be filled vs. no requiring anything.
- Columns (c) and (d) are analog to (a) and (b) but only truly correct series are considered.
- Note that $Z_2^\lambda$ does not explicitly appears in this table, it can be obtained by subtracting column (c) to column (a), or column (d) to column (d).
- All figures (including the *Ratios*) are averages across experiments. Then, figures in column *Ratios* are not necessarily equal to $c/a$ and $d/b$ because they are the mean across experiments.

that we are not always using the same penalty function in each experiment, but artificially helps the dynamic factors methodology to pick the correct number of factors (which is always one).

In order to have a proper comparison with the pairwise procedure, we compute confidence intervals for the factor loadings and identify those series with statistically significant factor loadings at significane level of 0.5%.[6] We consider the subset formed by those series as the *DFM counterpart of the single-cycle subset*. Using those series, we can compute the gauge and

[6] We use the asymptotic variance of the factor loadings provided by Bai (2003).




- Number of experiments: 500.
- $Z_2$ = number of wrong series included in $\hat{SC}$
- $Z_1$ = number of correct series included in $\hat{SC}$

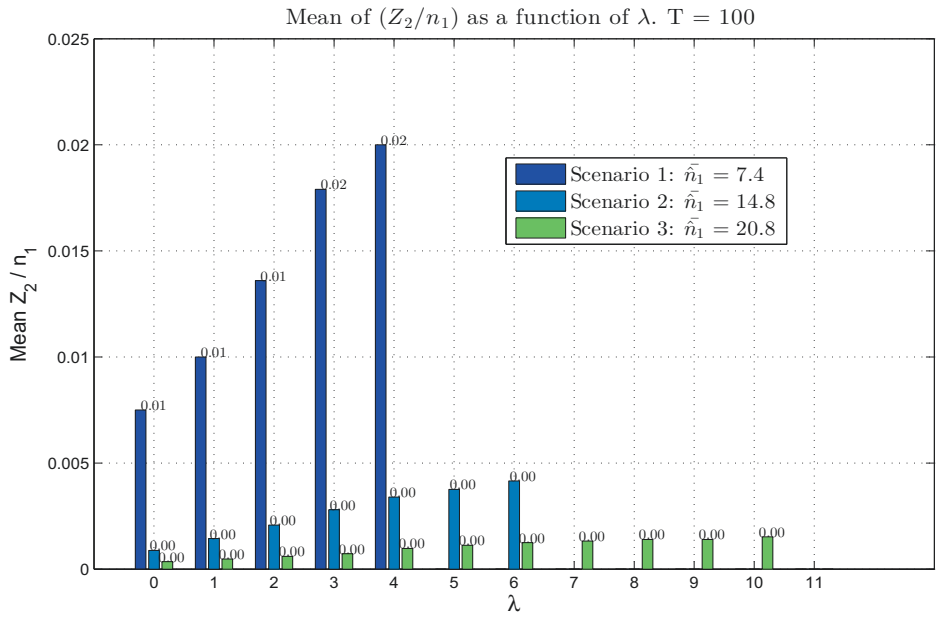
Figure IV.3: Mean of the ratio $Z_1/SC$ as a function of the relaxation parameter $\lambda$ ($\varphi = 0.05$)

potency of the DFM approach and the probability functions, $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$, that we used to study the performance of our approach.

For space reasons the probability density functions $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for the DFM procedure are not included. Instead, table IV.7 and table IV.8 include the *gauge* and *potency* of the pairwise strategy and its DFM counterpart for DGP 1 (VAR) and DGP 3 (DFM), respectively. Conclusions for DGP 2 are the same as those for DGP 1, so details are omitted.

Focus first on table IV.7 (DGP is a VAR). It shows that DFM procedure performs better than the pairwise approach in terms of *potency*. Although the latter approach shows quite high potencies for all scenarios and sample sizes, the probability of including a large proportion of the true series is larger when the DFM procedure is used. However, it is also the case that in the DFM approach the gauge is substantially increased in scenario 1, for any sample size, and in scenario 2 for $T = 100, 200$. As expected from §IV.4.3, the gauges of the pairwise procedure are close to zero for all scenarios and sample sizes.

Interestingly, for relatively large $SC$ (scenario 3) and $T > 100$, the DFM approach shows a *perfect* performance; it includes all the true series and almost none wrong ones. This outcome is in line with our analysis of §IV.3.2.3. There, we showed that a stationary VAR with *SCCF* is equivalent to a DFM with no idiosyncratic term. Using that result, it is easy to see that if the *SCCF* is not present in all the series, idiosyncratic components will appear in the series that do not have the feature. In such a case, the factor will explain *all* the variance of the series in $SC$




- Number of experiments: 500.
- $Z_2$ = number of wrong series included in $\hat{SC}$
- $Z_1$ = number of correct series included in $\hat{SC}$

Figure IV.4: Mean of the ratio $Z_2/SC$ as a function of the relaxation parameter $\lambda$ . $T = 100$ ($\varphi = 0.05$)

and the significant factor loadings will be very easily distinguishable from the non-significant ones. However, as showed in §IV.3.2.4, in the more general case that stationary commonalities are *WF*, idiosyncrasies will appear in all series and the DFM's performance will deteriorate (we analyze this issue in §IV.5.2.3)

We now turn to table IV.8, that includes the results for DGP 3 (DFM). In this case, the reduced form is a VMA with a reduced rank structure (see §IV.3.3.1). Given the DGP 3 described in §IV.5.1.1, series that do not have the factor are white noise so that, in a pair in which none of the series have the factor, there are two 'common' cycles. Thus, asymptotic power equal to one does not ensure anymore that wrong elements will not be included in $\hat{SC}$. Likewise, for a pair in which one series has the cycle and the other does not, there is one 'common' cycle. Therefore, we are in a situation in which outsiders can be included in $\hat{SC}$ because tests with insiders will deliver $s = 1$ (when not rejecting the true null hypothesis of $s > 0$ against $s = 0$), and tests with wrongly included outsiders can also deliver $s = 1$ (when wrongly rejecting the null $s > 1$ vs. $s = 1$, after having correctly not rejected $s > 0$ vs. $s = 0$).[7] Hence, to control the inclusion of wrong series, we should use tighter nominal sizes (compare gauges of the pairwise

[7] The first issue can be controlled by testing the significance of the common cycle' coefficients, and this is what we do in the empirical application of chapter VI.

Table IV.7: Comparison of Gauge and Potency of the Pairwise procedure with DFM. *Almost Full connection.* DGP 1, $\varphi = 0.05$

| | **Pairwise** (*almost full connection*) | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 0.2 | 88.2 | 0.2 | 83.1 | 0.1 | 79.1 |
| T=200 | 0.0 | 97.2 | 0.0 | 94.0 | 0.0 | 92.2 |
| T=400 | 0.0 | 97.9 | 0.0 | 95.4 | 0.0 | 93.9 |
| | **DFM counterpart** | | | | | |
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 30.1 | 94.0 | 7.2 | 99.9 | 1.4 | 99.3 |
| T=200 | 21.5 | 99.8 | 1.2 | 100.0 | 0.0 | 100.0 |
| T=400 | 10.4 | 100.0 | 0.1 | 100.0 | 0.0 | 100.0 |

- Number of experiments: 500.
- For scenarios 1 to 3, $SC$ is 10, 25 and 40 respectively.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- $Z_2$ = number of wrong series included in $\hat{SC}$
- $Z_1$ = number of correct series included in $\hat{SC}$
- $Nexp$ = number of experiments

procedure for nominal sizes 5% and 1% in table IV.8).

The increased risk of including wrong series is reflected in table IV.8, in which gauges of the pairwise procedure are notably larger than for DGP 1. Still, in scenario 1, the pairwise procedure beats its DFM counterpart, as the latter includes too many wrong series. For the other scenarios DFM outperforms the pairwise approach.

In regards to the pairwise procedure's potency, results clearly deteriorates with respect to DGP 1. The main reason for this result is that, as discussed in §IV.3.3.1, the reduced form of DGP 3 is a VMA with common cycles, not a VAR, but we are trying discover the common cycles in parsimonious VAR models.

In summary, the DFM procedure completely fails when $SC$ is small (scenario 1), regardless the DGP and sample size. For larger $SC$ its performance remarkably improves. When the DGP is a VAR with SCCF the DFM procedure shows a *perfect* performance for scenarios 2 and 3, what is explained by the DFM that corresponds to the DGP. In contrast, for this DGP the pairwise approach shows a very good performance for all scenarios and samples sizes.

When the DGP is a DFM, the DFM approach beats the pairwise, except for scenario 1.

Table IV.8: Comparison of Gauge and Potency of the Pairwise procedure with DFM. *Almost full connection.* DGP 3 (DFM)

| | **Pairwise** $\lambda = 2$, $\varphi = 0.05$ | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 3.5 | 51.6 | 2.7 | 49.6 | 1.9 | 48.1 |
| T=200 | 3.6 | 63.9 | 2.4 | 61.8 | 1.4 | 59.2 |
| T=400 | 3.7 | 72.3 | 2.5 | 70.5 | 1.3 | 68.2 |

| | **DFM counterpart** | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 34.2 | 95.8 | 3.7 | 87.0 | 0.2 | 74.6 |
| T=200 | 22.1 | 98.0 | 0.5 | 88.2 | 0.0 | 74.5 |
| T=400 | 11.5 | 99.1 | 0.0 | 88.7 | 0.0 | 74.4 |

| | **Pairwise**, $\lambda = 2$, $\varphi = 0.01$ | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 2.4 | 45.8 | 2.1 | 44.8 | 1.6 | 43.7 |
| T=200 | 2.6 | 61.4 | 1.6 | 60.1 | 0.9 | 57.6 |
| T=400 | 2.8 | 72.4 | 1.5 | 70.8 | 0.6 | 68.4 |

- Number of experiments: 500.
- For scenarios 1 to 3, $SC$ is $10, 25$ and $40$ respectively.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- $Z_2$ = number of wrong series included in $\hat{SC}$
- $Z_1$ = number of correct series included in $\hat{SC}$
- $Nexp$ = number of experiments

### IV.5.2 Non Stationary models

In this subsection we study the behavior of the pairwise approach applied to common cycles in a non-stationary framework. The analysis is composed by three parts; in §IV.5.2.1 we describe the DGPs, in §IV.5.2.2 we study the performance of the pairwise approach, and §IV.5.2.3 we compare our procedure with a DFM alternative.

#### IV.5.2.1 Monte Carlo design for non-stationary models

We want to simulate a situation in which a subset $CT$ and a subset $SC$ of series share a single common trend and a single common cycle, respectively (we will use $CT$ and $SC$ both, as the names of the subsets and to indicate the number of series inside them). Specifically, we want the data to satisfy the following conditions:

(i) $CT \cap SC$ is not empty;

(ii) $(CT \cap SC) \neq CT$ and $(CT \cap SC) \neq SC$ (neither of the two subsets contains the other);

(iii) $CT \cup SC \neq X_t$ (there are some series belonging neither to $CT$ nor to $SC$).

In order to simplify the system's dynamics we assume that the series not belonging to $CT$ or $SC$ have zero covariance with all the other series. As in §IV.5.1 we consider two DGPs; a VAR process and a DFM. We call the two DGPs as *Non-Stat-VAR* and *Non-Stat-DFM*, respectively.

**DGP: Non-Stat-VAR**

$$\Delta X_t = \alpha(\beta' X_{t-1} + c_0) + \Phi \Delta X_{t-1} + \epsilon_t, \tag{IV.62}$$

Matrices $\alpha$ and $\beta$ are those proposed in §II.3.1:

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_1 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_2 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_r \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{n \times r}, \beta' = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \cdots & & & & & & & \\ -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}_{r \times n},$$

where $r = CT - 1$; and $\alpha_i$ is taken from the uniform distribution with parameters $[0.15, 0.3]$. See §II.3.1 for a detailed discussion about the dynamic properties implied by these matrices. Then, we have:

$$\alpha\beta' = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \alpha_1 & -\alpha_1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \alpha_2 & 0 & -\alpha_2 & 0 & \cdots & 0 & \cdots & 0 \\ \alpha_3 & 0 & 0 & -\alpha_3 & \cdots & 0 & \cdots & 0 \\ \cdots & & & & & & & \\ \alpha_r & 0 & 0 & 0 & \cdots & -\alpha_r & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & & \\ \cdots & & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}_{n \times n}. \tag{IV.63}$$

For matrix $\Phi$ we keep the basic design of §IV.4.4.1 but, in order to fulfill conditions (i) to

(iii), the position of the square sub-matrix inside $\Phi$ is modified as follows:

$$\Phi = \begin{bmatrix}
\phi_{11} & 0 & \dots & & & & & & & & \dots & 0 \\
0 & \phi_{22} & 0 & \dots & & & & & & & \dots & 0 \\
\vdots & & & & & & & & & & & \\
0 & 0 & \dots & \phi_{mm} & 0 & \dots & & & & & \dots & 0 \\
0 & 0 & \dots & 0 & \phi_{m+1,m+1} & \dots & \phi_{m+1,m+sc} & 0 & \dots & & \dots & 0 \\
\vdots & & & & & & & & & & & \\
0 & 0 & \dots & 0 & \phi_{m+sc,m+1} & \dots & \phi_{m+sc,m+sc} & 0 & \dots & & \dots & 0 \\
0 & \dots & & & & \dots & 0 & \phi_{m+sc+1,m+sc+1} & 0 & 0 & \dots & 0 \\
0 & \dots & & & & \dots & 0 & 0 & \phi_{m+sc+2,m+sc+2} & 0 & \dots & 0 \\
\vdots & & & & & & & & & & & \\
0 & \dots & & & & & & & & \dots & 0 & \phi_{n,n}
\end{bmatrix},$$

where $m$ is the number of series that belong to $CT$ but not to $SC$ ($m < CT$). In a more compact way, $\Phi$ can be represented as:

$$\Phi = \begin{bmatrix} S_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S_{33} \end{bmatrix} \tag{IV.64}$$

where $S_{11}$ is diagonal and has dimensions $m \times m$; and $S_{33}$ is also diagonal with dimensions $(n - m - SC) \times (n - m - SC)$. In both cases the values of the diagonal are taken from the uniform distribution with parameters $[0.4, 0.7]$. $S_{22}$ has dimensions $(SC \times SC)$, has rank 1, and is generated as the upper left sub-matrix of $\Pi$ used §IV.4.4.1. In short, $S_{22} = \delta_{\perp}^{*} \Psi^{*'}$, where $\delta_{\perp}^{*}$ and $\Psi^{*}$ are $SC \times 1$ vectors.

Partition the vector $X_t$ in three sub-vectors of size $m$, $SC$, and $N - m - SC$, respectively and call $X_t^{sc}$ to the second sub-vector. Then, the common cycle is $\Psi^{*'} X_{t-1}^{sc}$, and $\delta_{\perp}^{*}$ contains the coefficients of the common cycle in each of the $SC$ series of the second sub-vector.

To study the dynamic properties of the DGP eq. (IV.62) we rewrite it as a VAR in levels:

$$X_t = c + \Pi_1 X_{t-2} + \Pi_2 X_{t-2} + \epsilon_t, \tag{IV.65}$$

where $\Pi_1 = I + \alpha\beta' + \Phi$; and $\Pi_2 = -\Phi$. The roots of the DGP are given the solutions of $|I - \Pi_1 \lambda - \Pi_2 \lambda^2| = 0$.

Using eq. (IV.64), the definition of $\Pi_1$, and eq. (IV.63), we conclude that $\Pi_1$ has a $3 \times 3$ block structure similar to that of eq. (IV.64):

$$\Pi_1 = \begin{bmatrix} S_{11}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S_{22}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S_{33}^* \end{bmatrix}.$$

Since $CT > m$, matrix $S_{11}^*$ has dimension $CT \times CT$, and contains the first $CT$ rows and columns of $(I + \alpha\beta' + \Phi)$. Matrix $S_{22}^*$ is square and has dimension $m + SC - CT$, but since we will assume $CT = SC$, its dimensions are $m \times m$, and contains the last $m$ rows and columns of $S_{22}$. Matrix $S_{33}^*$ has dimensions $(N - m - SC) \times (N - m - SC)$ and is equal to $I + S_{33}$.

Therefore, since $\Pi_2 = -\Phi$, the roots of the DGP's characteristic polynomial roots are the solutions of:

$$det \begin{pmatrix} I_m - S_{11}^*\lambda + \Phi_{[1:ct]}\lambda^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{sc} - S_{22}^*\lambda + \Phi_{[ct+1:ct+m]}\lambda^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{(n-m-sc)} - S_{33}^*\lambda + S_{33}\lambda^2 \end{pmatrix} = 0, \tag{IV.66}$$

where $\Phi_{[i:j]}$ stands for the sub-matrix that contains rows and columns $i$ to $j$ of matrix $\Phi$.

Using the result that the determinant of a block diagonal matrix is the product of the determinants of the sub-matrices in the diagonal, the solutions of eq. (IV.66) are those of:

a) $|I_m - S_{11}^*\lambda + \Phi_{[1:ct]}\lambda^2| = 0$;

b) $|I_{SC} - S_{22}^*\lambda + \Phi_{[ct+1:ct+m]}\lambda^2| = 0$; and

c) $|I_{(n-m-SC)} - S_{33}^*\lambda + S_{33}\lambda^2| = 0$.

In case (a), we will have the roots of a VECM with one lag and $CT$ series that share a unique common trend. Then, we need all the roots outside the unit circle except for one of them, that should be equal to one. In the case (b), we will have the roots of a VAR in levels with $m$ $I(1)$ series. Thus, we need to add $m$ unit roots. In case (c) we will also have the roots of a VAR in levels and we need $(N - CT - m)$ unit roots (recall that $SC = CT$).

In summary, we need $N - (CT - 1)$ unit roots and $CT - 1$ roots outside the unit circle. In contrast to §IV.4.4.1, a general way choosing the coefficients is not possible, so we proceed

as follows. For the diagonal matrices $S_{11}$ and $S_{33}$ coefficients are taken from the uniform distribution with parameters $[0.4, 0.7]$, and matrix $S_{22}$ is selected as in DGP 1 of §IV.4.4.1. Since these selections do not ensure $CT-1$ roots outside the unit circle, in each replication we compute the roots of the model, order them from the largest to the smallest, and disregard the model if the root in position $N-(CT-1)+1$ is larger than 0.95 in absolute value.

**DGP: Non-Stat-DFM**

We turn now to the design of the DGP when it is a DFM. We consider the following model:

$$\begin{aligned} X_t &= \Lambda F_{1t} + \Theta F_{2t} + \epsilon_t, \\ F_{1t} &= F_{1,t-1} + \eta_{1t}, \\ F_{2t} &= \rho F_{2,t-1} + \eta_{2t}, \end{aligned} \tag{IV.67}$$

where $\Lambda$ and $\Theta$ are a $N \times 1$ vectors of factor loadings, which are partitioned as follows: $\Lambda = [\Lambda'_{1_{(1\times m)}}, \Lambda'_{2_{[1\times(CT-m)]}}, \Lambda'_{3_{[1\times(N-CT)]}}]'$, and $\Theta = [\Theta'_{1_{(1\times m)}}, \Theta'_{2_{[1\times(CT-m)]}}, \Theta'_{3_{[1\times(SC-(CT-m)]}}, \Theta'_{4_{[1\times(N-SC)]}}]'$. The $m$ elements of $\Lambda_1$ and the $(CT-m)$ elements of $\Lambda_2$ are taken from the uniform distribution with parameters $[0.1, 0.8]$. The $(N-CT)$ entries of $\Lambda_3$ are zero. The $m$ elements of $\Theta_1$ are zero. The $CT-m$ entries of $\Theta_2$ are taken from the uniform distribution with parameters $[0.1, \sqrt{(1-\Lambda_{2i}^2)/2\sigma_{F_2}^2})$, as will became clear below, this upper limit ensures that $Var(\Delta X_t)$ can be 1 (without requiring negative idiosyncratic variance). The $SC-(CT-m)$ entries of $\Theta_3$ are taken from the uniform distribution with parameters $[0.1, 0.8]$, and the $(n-SC)$ values of $\Theta_4$ are zero.

The parameter $\rho$ is 0.6. The noise of the factors, $[\eta_1', \eta_2']' = \eta_t$, is distributed as $N(0, I_N)$, and $\epsilon_t$ is a $N \times 1$ vector, which is partitioned as:

$\epsilon_t = [\epsilon'_{t,1_{(1\times m)}}, \epsilon'_{t,2_{[1\times(CT-m)]}}, \sum_{i=1}^t \epsilon'_{i,3_{[1\times(SC-(CT-m)]}}, \sum_{i=1}^t \epsilon'_{i,4_{[1\times(N-SC)]}}]'$.

The vector $\epsilon_t^* = [\epsilon'_{t,1}, \epsilon'_{t,2}, \epsilon'_{t,3}, \epsilon'_{t,4}]'$ is distributed as an *iid* $N(0, \Sigma)$ process independent of $\eta_t$. For $i \leq m$, we set $\sigma_{ii}^2 = (1-\Lambda_{1i}^2)/2$. For $m < i \leq CT$, $\sigma_{ii}^2 = (1-\gamma_{2i}^2 - 2\theta_{2i}^2\sigma_{F_2}^2)/2$. Next, for $CT < i \leq SC+m$, $\sigma_{ii}^2 = 1-2\theta_{3i}^2\sigma_{F_2}^2$. Finally, for $i > SC+m$, $\sigma_{ii}^2 = 1$. These choices ensure $Var(\Delta X_{it}) = 1$ for all $i$.

Note that for $m < i \leq CT$, we need to impose some restrictions on the parameters in order to ensure $\sigma_{ii}^2 > 0$. The condition for positive variance is: $\theta_{2i} < \sqrt{\frac{1-\Lambda_{2i}^2}{2\sigma_{F_2}^2}}$.

With those variances, the *noise/signal* ratios are: $\frac{1-\Lambda_{1i}^2}{\Lambda_{1i}^2}$, for $i \leq m$; $\frac{1}{\Lambda_{2i}^2+2\theta_{2i}^2\sigma_{F_2}^2} - 1$, for

$m < i \leq CT$; $\frac{1}{2\theta_{3i}^2 \sigma_{F_2}^2} - 1$, for $CT < i \leq SC + m$; and 1 for $i > SC + m$.

**Scenarios**

For the two DGPs described above (*Non-Stat-VAR* and *Non-Stat-DFM*) we consider three scenarios. Using the notation $|A|$ to indicate the cardinality of set $A$, the three scenarios are:

**Scenario 1** : $|SC| = |CT| = 10$, and $|SC \cap CT| = 5$.

**Scenario 2** : $|SC| = |CT| = 25$, and $|SC \cap CT| = 12$.

**Scenario 3** : $|SC| = |CT| = 40$, and $|SC \cap CT| = 20$.

For each of these scenarios we consider three different sample sizes: $T = 100$, $T = 200$, and $T = 400$.

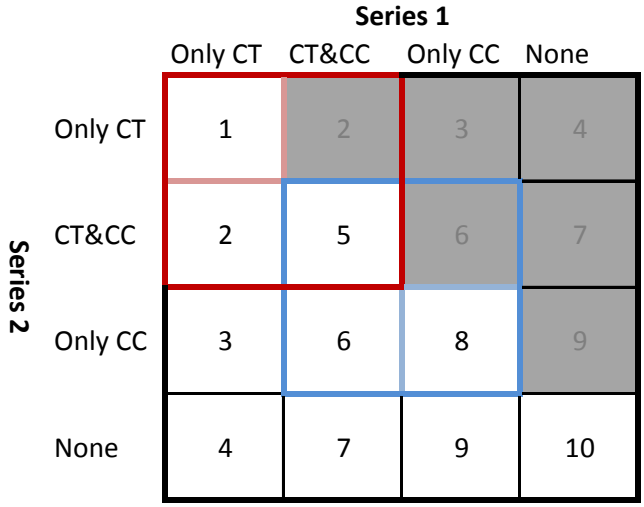
For each DGP, scenario and sample size, we perform 500 Monte Carlo replications. In each replication we simulate a 100-dimensional model in which a subset of $CT$ and a subset of $SC$ series share a single common trend and a single common cycle, respectively. The objective is to discover the series in $CT$ and $SC$. To do that, we first perform cointegration tests on all the 4950 bi-variate VAR sub-models that exist among the 100 series. Thus, for a particular DGP, scenario and sample size, we have 2.475 million sub-models (4950 for each replication). Since we have two DGPs, three scenarios and three sample sizes, we have $(2 \times 3 \times 3) \times 2.475 = 44.55$ million sub-models to estimate and in which we have to test for cointegration. Additionally, since the lag length for each of the 4950 sub-models of a particular replication is unknown, we select it with the AIC in a model with one cointegration relationship and admitting between one and five lags in the VEqM representation.

Once the results on cointegration are obtained we perform pairwise WF tests conditional on those results, thus we need to add other 44.55 million tests. In this case the lag length is unknown again so we estimate it in the same fashion as for cointegration tests, but using the estimated cointegration rank. The high complexity of the simulations led us not to consider other DGPs.

#### IV.5.2.2 The behavior of the pairwise strategy under non-stationarity

Assessing the performance of the pairwise strategy in this case is somewhat more tricky than in previous cases. There are three main reasons for that. First, we should analyze the ability

of the procedure to identify five different subsets of series (instead of only one as was the case of §IV.5.1 and chapter II). Figure IV.5 schematizes the data structure detailed in §IV.5.2.1 and highlights the five subsets of interests. The figure represents all the possible pairs of series. In each pair inside the dark red box (boxes 1, 2 and 5), both series of each pair have the common trend (*CT* subset); in the pairs inside the dark blue box (boxes 5, 6 and 8), both series have the common cycle (*SC* subset). While in box 5 all the series share the trend *and* the cycle (subset *Intersect*); in boxes 1 and 8 the series have only the common trend and only the common cycle, respectively (subsets *Only CT* and *Only SC*). In summary, the five different subsets are: the dark red box; the dark blue box; and boxes 1, 5, and 8.




- The red box represents series shearing a single common trend.
- The blue box represents series sharing a single common cycle (WF).
- Box 1: both series of each pair have only the common trend.
- Boxes 2: one series of each pair has only the common trend and the other has both, the common trend and the common cycle.
- Boxes 3: in each pair, one series has only the common trend and the other has only the common cycle.
- Boxes 4: both series of each pair have neither the common trend, nor the common cycle.
- Box 5: both series of each pair have the common trend *and* the common cycle.
- Boxes 6: one series of each pair has only the common cycle and the other has both, the common trend and the common cycle.
- Boxes 7: One series of each pair has both the common trend and the common cycle, and the other has none of them.
- Box 8: Both series of each pairs have only the common cycle.
- Boxes 9: One series of each pair has only the common cycle, and the other has neither the common trend, nor the common cycle.
- Box 10: Both series of each pair have neither the common trend, nor the common cycle.

Figure IV.5: Different types of pairs according to the presence of common features in each series

Second, the relaxation to *almost full connection* in this case is more complex than in previous exercises, for trying different $\lambda$s for $\hat{CT}$ and $\hat{SC}$ would lead to a quite large number of possible combinations. To avoid this added complexity, we fix the $\lambda$ for relaxing the cointegration tests ($\lambda_{CT}$), and to perform the WF test we use the cointegration relationships coming form the relaxed $\hat{CT}$.

This simplifying decision is based on two arguments. First, as showed in chapter III, the cost of relaxing the strict full cointegration requirement is small. Second, as argued by Hecq et al. (2006) the cost of correcting WF tests with $I(1)$ variables, that should not be considered, is almost zero. Therefore, even if we could include a few wrong series in $\hat{CT}$ this will not affect WF tests.

We set $\lambda_{CT} = min\{6, 0.4 \times \hat{CT}^*\}$, where $\hat{CT}^*$ represents the number of series in the estimated *strict* fully cointegrated set. Though in chapter III we argued that the relaxation was required only for $T = 100$, for a reason that will become clear below, we apply it for all sample sizes and scenarios.

The third reason that makes the evaluation of the pairwise strategy more complex in this case is that the cointegration relationships to correct the common cycles tests can be used in five different ways (for a discussion about the inclusion of cointegration relationships in common cycles tests see §IV.4.2):

(i) Use all cointegration relationships, tightening significance levels for pairs outside the estimated dark red box (e.g., 1/1000 or 5/1000).

(ii) Use all cointegration relationships inside the estimated dark red box, but only those.

(iii) Use only the cointegration relationship of the pair.

(iv) For pairs inside the estimated dark red box use only the cointegration relationships that have non-zero coefficients in each series.

(v) Do not use cointegration relationships at all.

Note that in options (i) and (ii) the inclusion of *all* the cointegration relationships inside the dark red box could lead to include irrelevant cointegration relationships for some pairs, as not all coefficients of matrix $\alpha$ are necessarily different form zero. Options (iii) and (iv) are designed to consider this issue.

In in what remains of §IV.5.2.2 we analyze the performance of the pairwise procedure for the five options. We do this only for the DGP *Non-Stat-VAR*. The best option will be the one to be used in the comparison of our procedure with the DFM alternative.

For space reasons we do not include the details for all the five possibilities, but focus on options (i), (iv) and (v) for the almost fully connected SC subsets. Results for option (ii)

are exactly the same as those for option (i), and results for option (iii) are an intermediate situation between (i) and (iv). Tables IV.9 to IV.11 contain the results for options (i), (iv), and (v), respectively.

The tables include the gauge and potencies for the five groups highlighted in figure IV.5, i.e., *CT*, *OnlyCT*, *CC*, *OnlyCC*, *Intersect*, for different sample sizes ($T = 100$, $T = 200$, and $T = 400$) and scenarios.

**Cointegration tests**

The five options do not affect cointegration results, and figures under *CT* columns are the same for the three tables. For large samples ($T = 400$), results are similar to those in chapter II. This was expected, as the reduced rank of $S_{22}$ (see eq. (IV.64)) is asymptotically irrelevant for cointegration tests. The short samples deterioration is now more notable than in chapter II, for the reduced rank of $S_{22}$ is not irrelevant in short samples. To see the importance of this issue compare the potencies under *CT* and *OnlyCT* columns; they are systematically larger in the second case (for $T = 400$ differences are minor). This is the reason for using the relaxation procedure for all sample sizes. Note that comparing gauges of these two columns is non-sense since they depend on the results of the common cycles tests.

**WF tests**

Consider now common cycles tests. For large samples ($T = 400$) we found that options (i) and (ii) — which gave the same results — largely outperform (iii) and (v) in terms of potency, compare *CC* columns of table IV.9 and table IV.11. In terms of gauge, all possibilities perform equally well. This conclusion highlights the importance of correcting the common cycles tests with cointegration relationships, and is in line with results in Hecq et al. (2006). When the true cointegration rank is $r^*$, the authors show that while specifying $r > r^*$ is almost costless for WF tests, using $r < r^*$ is highly damaging.

For $T = 200$, the dominance of (i) and (ii) with respect to (iii) and (v) is less strong for scenarios 1 and 2, and reverted for scenario 3. For $T = 100$, the dominance is reverted for scenarios 2 and 3. As scenarios 2 and 3 are those with larger *CT*, this result highlights the curse of dimensionality issue in short samples. It is important to remark that this result should not be compared with that of Hecq et al. (2006) about the little cost of over-specifying the cointegration rank. There are two reasons that invalidate the comparison. First, since the

'extra' cointegration relationships do not exist, in the case of Hecq et al. (2006), the over-specification leads to include an 'extra' I(1) process in the tests. In contrast, in our case, the 'extra' cointegration relationships are I(0) processes with non-significant coefficients. Second, Hecq et al. (2006) work with low dimensional problems and the curse of dimensionality is not relevant.

In option (iv) we include only the relevant cointegration relationships for each series of the pair. In a fully cointegrated subset of $CT$ series there are $(CT-1)$ cointegration relationships, but not all the elements of matrix $\alpha$ are necessarily different from zero. In fact, in our simulation design, matrix $\alpha$ has many zeros and each series (except for the first one) reacts only to one cointegration relationship. Therefore, in option (i) we are including a large amount of noise in the *WF* tests. For example, in scenario 3 we are considering up to thirty nine cointegration relationships when we should include only two of them in each pairwise test. This observation explains the dominance of option (v) over (i) in scenarios 2 and 3 for $T = 100$.

When only the truly relevant cointegration relationships are considered for the *WF* tests, results are better than in option (i) and (v) for all scenarios and sample sizes. Compare potencies under *CC* columns of table IV.10 with those of tables IV.9 and IV.11. The improvement with respect to option (i) is quite important for scenarios 2 and 3, in which the 'added noise' in that option is larger.

It is important to note that in option (iv) we are not statistically selecting the relevant cointegration relationship, but using our knowledge of the DGP. In empirical applications, selection must be carried out, and results of table IV.10 may deteriorate. Additionally, for the empirical applications, we will consider cointegration relationships outside the almost fully cointegrated subsets only if they appear when tighter significance levels are used.

Finally, compare table IV.10 with the first block of table IV.7 (stationary VAR with *SCCF*). Though of the same order of magnitude, results are somewhat better in the stationary case. Conclusions extracted for the stationary case are, basically, still valid. The slight difference between the two cases is due to, at least, two reasons: first, we are using estimated cointegration relationships; second, the lag length is larger in the non-stationary case for the VECM(1) corresponds to a VAR(2).

In summary, from tables IV.9 to IV.11 we extracted two important conclusions: (1) We confirmed the importance of correcting the WF tests highlighted by Hecq et al. (2006). (2) We

found that the potency of the pairwise approach to discover the single-cycle subsets could be severely affected if the cointegration relationships are not properly selected. In particular, when matrix $\alpha$ has zero restrictions, including only those cointegration relationships with significant coefficients in matrix $\alpha$ could lead to remarkable potency improvements. This is because a smaller number of variables reduces the estimation uncertainty.

Table IV.9: Gauge and Potency of the Pairwise approach. WF tests with all cointegration relationships. *Almost Full connection*. DGP: 'Non-Stat-VAR'

| | **Pairwise** (All CI rels -option (i) at 1/1000- ) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.9 | 97.8 | 0.0 | 95.2 | 0.9 | 98.4 | 0.0 | 93.7 | 0.1 | 94.3 |
| Sce 2 | 0.8 | 97.1 | 0.0 | 87.4 | 1.8 | 97.9 | 0.1 | 85.5 | 0.2 | 87.5 |
| Sce 3 | 0.7 | 97.9 | 0.0 | 72.1 | 7.3 | 98.5 | 0.0 | 72.6 | 0.1 | 70.9 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.6 | 88.6 | 0.0 | 87.0 | 0.9 | 95.9 | 0.5 | 90.5 | 0.1 | 74.1 |
| Sce 2 | 0.6 | 85.0 | 0.2 | 71.3 | 3.4 | 96.3 | 1.4 | 82.1 | 0.3 | 53.0 |
| Sce 3 | 0.7 | 81.5 | 1.1 | 56.4 | 13.6 | 92.8 | 2.6 | 87.1 | 0.9 | 14.8 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.8 | 44.9 | 1.0 | 80.3 | 1.0 | 58.3 | 3.7 | 79.3 | 0.4 | 22.2 |
| Sce 2 | 0.8 | 49.9 | 1.6 | 69.5 | 3.0 | 69.8 | 8.3 | 75.3 | 1.0 | 10.9 |
| Sce 3 | 1.0 | 47.5 | 2.9 | 64.5 | 5.8 | 65.7 | 12.5 | 77.4 | 1.8 | 2.8 |

- Number of experiments: 500.
- For scenarios 1 to 3, $SC$ is $10, 25$ and $40$ respectively.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- $Z_2$ = number of wrong series included in the corresponding subset.
- $Z_1$ = number of correct series included in the corresponding subset.
- $Nexp$ = number of experiments
- *CT* column represents the dark red box of figure IV.5.
- *CC* column represents the dark blue box of figure IV.5.
- *OnlyCT* column represents box 1 of figure IV.5.
- *OnlyCC* column represents box 8 of figure IV.5.
- *Intersect* column represents box 5 of figure IV.5.

### IV.5.2.3 Comparison with DFM under non-stationarity

We now compare the pairwise strategy with a DFM counterpart. For space reasons we focus on the DGP '*Non-Stat-VAR*'. The results for the DGP '*Non-Stat-DFM*' are included in ap-

Table IV.10: Gauge and Potency of the Pairwise approach. WF tests with the truly relevant cointegration relationships. *Almost Full connection*. DGP: 'Non-Stat-VAR'

| | **Pairwise** (Only relevant CI rels - option (iv)-) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.9 | 97.8 | 0.0 | 95.2 | 0.8 | 98.4 | 0.0 | 93.4 | 0.1 | 94.6 |
| Sce 2 | 0.8 | 97.1 | 0.0 | 92.3 | 1.2 | 97.9 | 0.0 | 91.6 | 0.2 | 91.3 |
| Sce 3 | 0.7 | 97.9 | 0.0 | 91.1 | 2.0 | 98.6 | 0.0 | 90.6 | 0.2 | 90.6 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.6 | 88.6 | 0.0 | 89.5 | 0.7 | 96.4 | 0.4 | 92.8 | 0.1 | 77.1 |
| Sce 2 | 0.6 | 85.0 | 0.0 | 83.4 | 1.1 | 97.7 | 1.1 | 91.3 | 0.1 | 67.9 |
| Sce 3 | 0.7 | 81.5 | 0.0 | 80.4 | 1.8 | 95.8 | 2.1 | 90.6 | 0.3 | 60.4 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.8 | 44.9 | 0.7 | 83.0 | 0.8 | 61.1 | 3.6 | 81.7 | 0.2 | 25.1 |
| Sce 2 | 0.8 | 49.9 | 0.6 | 76.4 | 1.2 | 74.8 | 8.0 | 78.3 | 0.4 | 21.3 |
| Sce 3 | 1.0 | 47.5 | 0.5 | 72.9 | 1.5 | 72.1 | 12.2 | 78.4 | 0.5 | 18.2 |

See notes to table IV.9.

pendix C.2. As in chapter II, the DFM counterpart consists of estimating a DFM for all the N series, selecting the number of factors, and keeping those series with statistically significant factor loadings (see §II.3.4 for the details). For this procedure to succeed, it should find two factors, one $I(1)$ and one $I(0)$, the former generated by the series in $CT$ and the latter by those in $SC$.

In §IV.3.2.4 we argued that when the DGP is a cointegrated VAR with non-pervasive common trends, the DFM should be estimated in the differenced data to avoid the presence of $I(1)$ idiosyncrasies. Hence, as in chapter II, we estimate the DFM in differences.

In §IV.3.2.4 we also argued that the first difference of an $I(1)$ time series coming from a cointegrated VAR with WF, can be written as a DFM either with $n-r$ common trends, or with $r+(n-s)$ common stationary factors. Which of the two representations will show up in the data depends on which factor structure represents a higher proportion of the variance, so it is an empirical matter.

In the DGP '*Non-Stat-VAR*', the $r$ cointegration relationships are in fact idiosyncratic, for they affect only one variable each. Thus, when estimating a DFM in the differences of the series

Table IV.11: Gauge and Potency of the Pairwise approach. SCCF tests. *Almost Full connection*. DGP: 'Non-Stat-VAR'

| | **Pairwise** (SCCF test -option (v)-) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.9 | 97.8 | 0.0 | 48.6 | 5.1 | 98.3 | 0.0 | 81.2 | 0.1 | 14.9 |
| Sce 2 | 0.8 | 97.1 | 0.0 | 50.4 | 13.5 | 97.9 | 0.0 | 94.5 | 0.2 | 8.2 |
| Sce 3 | 0.7 | 97.9 | 0.0 | 51.0 | 23.1 | 98.6 | 0.0 | 94.7 | 0.2 | 6.2 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.6 | 88.6 | 0.1 | 60.3 | 2.9 | 95.8 | 0.6 | 75.4 | 0.1 | 35.7 |
| Sce 2 | 0.6 | 85.0 | 0.0 | 59.5 | 7.8 | 97.6 | 1.6 | 88.1 | 0.1 | 22.7 |
| Sce 3 | 0.7 | 81.5 | 0.0 | 60.3 | 12.2 | 95.8 | 2.6 | 89.6 | 0.3 | 19.4 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 0.8 | 44.9 | 0.7 | 79.8 | 1.0 | 60.3 | 3.6 | 78.2 | 0.2 | 22.4 |
| Sce 2 | 0.8 | 49.9 | 0.8 | 73.5 | 1.6 | 74.4 | 8.4 | 72.4 | 0.4 | 19.1 |
| Sce 3 | 1.0 | 47.5 | 0.8 | 70.9 | 2.2 | 71.8 | 12.9 | 74.3 | 0.6 | 15.9 |

See notes to table IV.9.

coming from that DGP, we expect to find either the common trend, or the common cycle.

We first study the performance of the DFM procedure when fixing the number of factors in two. Results are included in table IV.12. Focusing on $CT$ and $CC$ columns we conclude that, for $T = 400$, the procedure does a relatively good job for identifying the series in $SC$, but not in $CT$. For instance, the gauges and potencies in scenario 3 are, 8.6% and 95.8%; and 39% and 62.3%, for subsets $SC$ and $CT$, respectively. These results indicate that the procedure completely fails in discovering the subset $CT$. Indeed, the information criteria for selecting the number of factors suggested by Bai (2004) pick only one factor in almost all the experiments.

Given these outcomes, we repeat the procedure but keeping only one factor, table IV.13 contains the new results. In general, we get somehow smaller gauges but also smaller potencies than those in $CC$ columns of table IV.12.

For comparing the pairwise procedure with its DFM counterpart we focus just on the subset $SC$ as — for the DGP we are dealing with —, the DFM procedure fails in discovering $CT$ for all scenarios and sample sizes. Compare therefore tables IV.10 and IV.13. Except for scenario 3 with $T = 400$, the DFM procedure beats the pairwise in terms of potency. This larger potency

Table IV.12: Gauge and Potency of the DFM approach. Two factors. DGP: 'Non-Stat-VAR'

| | **DFM counterpart** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 55.8 | 59.5 | 9.1 | 99.6 | 55.3 | 73.8 | 13.9 | 99.3 | 9.5 | 45.1 |
| Sce 2 | 36.9 | 70.1 | 13.0 | 98.8 | 37.8 | 99.5 | 25.9 | 97.6 | 24.6 | 42.8 |
| Sce 3 | 39.0 | 62.3 | 8.6 | 95.8 | 35.9 | 98.1 | 31.5 | 91.6 | 29.1 | 26.6 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 55.0 | 51.7 | 16.0 | 98.9 | 54.4 | 60.1 | 20.4 | 97.9 | 11.7 | 43.3 |
| Sce 2 | 46.5 | 67.0 | 12.1 | 98.3 | 45.7 | 95.0 | 25.1 | 96.5 | 22.9 | 41.1 |
| Sce 3 | 45.9 | 62.2 | 9.5 | 94.7 | 41.7 | 95.4 | 32.2 | 89.4 | 28.9 | 29.0 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 52.0 | 41.9 | 25.6 | 97.5 | 51.5 | 42.1 | 29.5 | 96.0 | 15.0 | 41.0 |
| Sce 2 | 52.5 | 57.8 | 13.1 | 97.0 | 50.4 | 78.5 | 25.9 | 94.1 | 19.0 | 38.4 |
| Sce 3 | 53.4 | 61.6 | 11.4 | 93.7 | 48.5 | 89.2 | 33.5 | 87.6 | 28.4 | 33.9 |

comes at the cost of remarkably larger gauges. As usual, the DFM procedure completely fails when the number of series that share the factor is not large, as the large gauges indicate.

Given the very small gauges of the pairwise procedure (close to zero in all situations), and that its potencies, in spite of being smaller than those of the DFM approach, are quite large, we conclude conclude that this procedure is the preferred one.

Table IV.13: Gauge and Potency of the DFM approach for discovering the *SC* subset. One Factor. DGP: 'Non-Stat-VAR'

| | **DFM counterpart** | | | | | |
|---|---|---|---|---|---|---|
| | Sce 1 | | Sce 2 | | Sce 3 | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| T=100 | 23.9 | 96.8 | 11.7 | 95.5 | 7.7 | 87.7 |
| T=200 | 15.2 | 98.5 | 10.6 | 97.0 | 5.6 | 88.6 |
| T=400 | 8.9 | 99.4 | 11.1 | 96.7 | 4.4 | 89.6 |

See notes to table IV.9.

## IV.6 Preliminary discussion for case of $N$ going to infinity

Up to now, the asymptotic analysis was for the case of 'large' but fixed $N$, and $T$ going to infinity. As we argued in chapter II, we consider that being able to deal both with 'large' and small $N$ is a strength of our procedure as we do not need to assume that the number of components of an aggregate goes to infinity. However, in order to fully study the properties of our strategy, analyzing its behavior when $N \to \infty$ may be of interest. In this section we give some initial insights about this case.

We split the analysis in two parts; (i) the size of the single-cycle subsets ($sc$) is fixed, and (ii) it goes to infinity.

### IV.6.1 Fixed *sc*

If the true size of single-cycle subsets ($sc$) remains fixed as $N$ increases, the problem of including a large proportion of the true series inside the estimated single-cycle subset is exactly the same as that analyzed in §IV.4.3. Thus, we focus on the inclusion of wrong elements and the false discovery of single-cycle subsets.

For the inclusion of wrong elements, define $Z_i$ as the random variable that takes the value one if the variable $i$ is wrongly included and zero otherwise. Then, the number of wrong inclusions is:

$$\sum_{i=1}^{N-sc} Z_i. \tag{IV.68}$$

Therefore, the expected number of wrong inclusions is:

$$E[\textstyle\sum_{i=1}^{N-sc} Z_i] = \sum_{i=1}^{N-sc} E[Z_i] = (N - sc)E[Z_j] \tag{IV.69}$$

Since asymptotic power of the common cycle test is 1, using assumption 7, $E[Z_i] = 0$. Thus, the expected number of wrong elements (eq. (IV.69)) is zero. Note that this result does not depend on the degree of dependency between the $Z_i$'s.

Regarding the discovery of false single-cycle subsets, we can use the same argument as above but for the expected number of pairs outside the single-cycle subsets for which we will wrongly find a common cycle. Define $R_i$ as the random variable that takes the value 1 if a common

cycle is found for the pair $i$ and zero otherwise. Then, the number of pairs for which we will wrongly find a common cycle is:

$$\sum_{i=1}^{Q} R_i, \tag{IV.70}$$

where $Q$ is the total number of pairs between components outside the single-cycle subsets. Therefore, the expected number of wrong common cycles is:

$$E[\textstyle\sum_{i=1}^{Q} R_i] = \sum_{i=1}^{Q} E[R_i] = QE[R_j], \tag{IV.71}$$

Again, since asymptotic power of the common cycle test is 1, using assumption 7, $E[R_i] = 0$. Thus, the expected number of wrong elements (eq. (IV.71)) is zero. Again, we can deal with the case of $N/T \to \infty$.

Note that nothing in this argument requires $sc$ and $N$ to grow at the same rate, so we can deal with non-pervasive common cycles.

### IV.6.2 $sc \to \infty$

In this case we need to study the problem of including a large proportion of the true series in the estimated single-cycle subsets. If $sc \to \infty$, the analysis in §IV.4.3.2 applies. There, we argued that under independence (which is the worst possible case), the required value of $1 - \varphi$ (where $\varphi$ is the nominal size of each individual test) to find a subset of size $sc$ in which all the series showed a common cycle with $k$ of the others, is given by eq. (IV.59). Then, choosing $k = \delta(sc - 1)$ and substituting in eq. (IV.59) we get:

$$\frac{1}{[[(\delta(sc-1))-1]sc]^{\frac{1}{[\delta(sc-1)]-1}}}.$$

The expression above tends to 1 at the same rate as $sc^{\frac{-2}{sc}}$. Then, as $sc \to \infty$ we need to set the significance level proportional to $(1 - sc^{\frac{-2}{sc}})$. For being able to use such a significance level, we will need $T$ to diverge at a rate larger than or equal to $[1 - sc^{\frac{-2}{sc}}]^{-1}$. Therefore, we can also deal with the case of $N/T \to \infty$ when $sc \to \infty$.

## IV.7 Concluding remarks

In this chapter we started providing a review of the literature on common cyclical features, and analyzing the relationships between VAR models with common features restrictions and Dynamic Factor Models (DFM). The pairwise strategy for dealing with common cycles, particularly if the data also contains common trends, faces a much greater number of difficulties than in the case of only common trends. Therefore, the proposal of a pairwise strategy in this new context requires a careful study of how to formulate the bi-variate VAR models, and of the tests' properties carried out in those models. This chapter included those studies and also compared the performance of the pairwise strategy with a DFM alternative.

The study of common cyclical features has been receiving increasing attention since Engle and Kozicki's (1993) seminal paper. Important improvements in the definition and estimation and testing procedures have been developed until Cubadda's (2007) 'weak form polynomial serial correlation feature' which, so far, seems to be the more general formulation of 'common cycles'. The concepts and definitions have moved from simple and easily interpretable structures, to more complex and economically diffuse ones. The Engle and Kozicki's (1993) original definition has the evident advantage of its economic interpretability in terms of cycles, but has the drawback of being too rigid to describe a wide range of possible short-run commonalities. On the other hand, the more complex concepts allow for richer structures between time series data, at the cost of less clear economic interpretation.

Although the only definition of common cyclical features that is equivalent to common cycles in the Beverdige-Nelson sense is the original *serial correlation common feature* (SCCF), proposed by Engle and Kozicki (1993), we used the term 'common cycle' in a broad sense to refer to any kind of common cyclical feature.

Regarding the relationship between restricted VAR models and DFM, we split the analysis into two parts. In the first one, we considered a restricted VAR as the original DGP and derived the equivalent DFM. In the second one, the DGP was a DFM and we derived the corresponding VARMA. The main conclusion was that, when the original DGP has both I(0) and I(1) common features (factors), there is not always a direct relationship between restricted VARs and DFM. In particular, when the DGP is a VAR (DFM) with cointegration and *weak serial correlation common feature* (I(1) and I(0) factors) the equivalent DFM (VAR) will not necessarily show I(1) and I(0) factors (common trends and common cyclical features).

Our strategy for dealing with common cyclical features in the framework of the pairwise approach is similar to that for dealing with common trends; test for common cyclical features between all the possible $N(N-1)/2$ pairs that exist in a set of $N$ series and look for subsets in which all pairs showed a single common cycle. For this procedure to be sensible, the definition of common cycle has to be transitive. We need that in a group of three variables, the fact that two pairs of series — out of the possible three — have a common cycle, implies that the third pair also has it. We showed that although the only transitive definition is the original SCCF, the other definitions can also be transitive after a simple extension of the testing strategy.

Assuming that the components are generated by a (big) VAR, the extension of the testing strategy for making *weak serial correlation common features* (WF) to be transitive, consists of including in each pairwise test all the cointegration relationships of the full system that are relevant for the series of the pair. For doing this we proceed in two steps, first we perform pairwise cointegration tests and then we perform the pairwise WF tests using the cointegration relationships of the first step. This two steps strategy raises three issues: a) We are not considering cointegration relationships which are not detectable with the pairwise strategy. b) What nominal size should be used for the cointegration tests between series outside *fully cointegrated* subsets? c) For series inside some *fully cointegrated* subset, not all the cointegration relationships are necessarily relevant for both series, then, we could be introducing noise in the *WF* tests if cointegration relationships are not carefully selected.

In this chapter we studied these three issues and concluded that although the first one is not solvable within the pairwise strategy, it could generate only potency problems, but will not affect gauge. For the other two problems, we concluded that they are not relevant when the number of cointegration relationships is small. If the number is not small, cointegration relationships should be selected before the WF tests (e.g., with *Autometrics*).

For studying the properties of the pairwise strategy applied to common cycles, we focused on two issues. First, given the large amount of tests we are performing, we may be inflating the false rejection probability — multiple testing problem. Second, what are the costs of proceeding in a pairwise fashion in comparison with a full model strategy, when the number of variables is reduced and the latter procedure feasible?

Common cycles tests are tests for zero canonical correlations, and the usual practice is to perform sequential tests. In the bi-variate case, we first contrast the null of $s > 0$ *vs.* $s = 0$

and, if not rejected, we then test $s = 1$ *vs.* $s = 2$. Since the asymptotic power of the tests is equal to one, when none of the series are white noise, the only multiple testing issue that we may face is falsely rejecting $s = 1$ *vs.* $s = 2$. In terms of the subsets we are looking for, this means that we may find subsets with too few series, but not with incorrect elements.

In contrast with our results on chapter II, we showed that the multiple testing problem is indeed present in common cycles tests. The reason is that the fact that the series inside 'single-cycle' subsets share the 'cycle' does not ensure that the pairwise test statistics are going to be dependent between each other. In fact, we showed that, under some conditions, they can be independent. The absence of a high degree of dependency makes the probability of including a large proportion of the true series inside the estimated 'single-cycle' subset to be a decreasing function of the true size of the subset. This is an undesirable property for we want our procedure to work well not only when subsets are small, but also when they are large.

We showed that a simple 'relaxation' strategy solves the problem with almost zero cost. The strategy consists of including in the estimated 'single-cycle' subsets the series for which the tests showed a common cycle with *almost* all the series in the subset. That is, we are allowing some of the tests to fail in showing a single-cycle. We studied the properties of this strategy by Monte Carlo in two different frameworks. First, we considered a *pure and clean* framework, in which we simulated random graphs of different sizes and studied the performance of the relaxation procedure. The results showed remarkable improvements in the sizes of the estimated subsets. The other framework was a more realistic one, in which the data is generated by a VAR process with common cycles. The performance in this case is also very good, and the costs are almost zero.

Regarding the comparison with the full model strategy we performed a Monte Carlo study in which the DGP was a small dimensional VAR. The results showed that, when the common cycles are pairwise detectable, nothing is lost for proceeding by pairs. On the contrary, relevant power gains for discovering the true number of common cycles relationships can emerge from the pairwise procedure.

Finally, we compared the performance of the pairwise procedure with a DFM alternative, for stationary and non-stationary processes, and analyzed the results under the light of the aforementioned study of the relationship between restricted VAR models and DFM. The main conclusion was that, while the DFM fail to discover the 'single-cycle' subsets when their sizes

are not large, the pairwise approach shows a good performance in any situation.

# Appendix C

# Appendix to Chapter IV

## C.1 Dynamic Factor Models: methodological issues

The starting point of DFMs is that the dynamics of a high dimensional ($n$) time-series vector ($X_t$) are driven by few ($q$) common factors $f_{it}$ and an idiosyncratic $n$-vector of disturbances $e_t$. The use of DFMs in economics became widespread after Geweke (1977) and Sargent and Sims (1977) who allowed both, the factors and the idiosyncrasies to be serially correlated. The factors ($f_t$) are usually assumed to follow a VAR process whereas the idiosyncratic disturbances ($e_t$) are assumed to follow univariate autoregressive processes. Thus, DFMs can be written as:

$$\begin{aligned} X_t &= \lambda(L) f_t + e_t \\ \Gamma(L) f_t &= \eta_t, \end{aligned} \tag{C.1}$$

where, the lag polynomials $\lambda_i(L)$ are the *dynamic factor loadings* of each series in $X_t$. Assume initially that both lines in equation eq. (C.1) are stationary. The idiosyncratic error $e_t$ is assumed to be uncorrelated with factors' innovations at all leads and lags ($E(e_t, \eta'_{t-k}) = 0 \; \forall k$). In the *exact dynamic factor model* it is also assumed that idiosyncratic disturbances are mutually uncorrelated at all leads and lags, that is, $E(e_{it} e_{js} = 0) \; \forall s$ if $i \neq j$.

As noted by Stock and Watson (2011), when the factors are known and the errors ($e_t$ and $\eta_t$) Gaussian, an individual variable can be efficiently forecast regressing it on the lagged factors and lags of the variable itself, so that we do not need to include all the $n$ variables in the regression. Thus, in words of Stock and Watson (2011) DFMs allow to 'turn dimensionality from a curse into a blessing'. However, not only the factors are unknown but also we do not know how many of them are driving the data, so this efficiency may be lost. In next two sub-sections we review

the details for estimating these two unknowns.

### C.1.1 Factor Estimation

Denoting the $r \times 1$ vector $(f_t', ..., f_{t-p}')'$ as $F_t$ and the $n \times r$ matrix $(\lambda_0, ..., \lambda_p)$ as $\Lambda$, where $\lambda_i$ is the $n \times q$ matrix of coefficients on the $i^{th}$ lag in $\lambda(L)$, then the DFM can be re-written in its static form as:

$$\begin{aligned} X_t &= \Lambda F_t + e_t \\ A(L)F_t &= G\eta_t \end{aligned} \tag{C.2}$$

where $A(L)$ contains $1s$, $0s$ and elements of $\Gamma(L)$; and $G$ is composed of $1s$ and $0s$. Note that the number of static factors will be $r \leq pq$ because some lagged factors could be redundant. As will became evident below, this state-space formulation has important advantages for estimation.

As indicated by Stock and Watson (2011), estimation methods can be divided in three classes. The first class considers a small number of series so that factors and model's parameters can be estimated using the Gaussian maximum likelihood (MLE) and the Kalman filter. The second class of approaches are those using non-parametric estimation via some averaging method among which principal components is the most usual. Finally, as factors can be consistently estimated by principal components (for large $n$), in the last class of methods these estimations are used to estimate the parameters of the state-space model, solving the dimensionality issue of the first approaches.

#### C.1.1.1 MLE and the Kalman filter

Kalman filter estimation needs some assumption about idiosyncratic disturbances' dynamics so, as aforementioned its usually assumed that $e_{it}$ is a univariate autoregression:

$$d_i(L)e_{it} = \xi_{it}, \ i = 1, .., n, \tag{C.3}$$

where $\xi_{it}$ is i.i.d $N(0, \sigma^2_{\xi_i})$. The Gaussian likelihood of state-space model in eq. (C.2) and eq. (C.3) can be constructed via the Kalman filter, and MLE estimators of the parameters obtained maximizing it. Given some initial conditions, the normality assumption for $\eta_{it}$ and $\xi_{it}$ ensures that the filtered estimates delivers the minimum mean squared error (MSE) estimator of the factors. When the disturbances are not normal, the estimates are minimum MSE among linear estimators. However the factors are usually estimated using some smoothing algorithm

since they use the full data up to the last observation $T$.

### C.1.1.2 Principal Components

As noted by Stock and Watson (2011) an important motivation of estimation approaches based on principal components is that in a (weighted) cross-sectional average of $X_t$, idiosyncratic disturbances will converge to zero so that only the linear combinations of the factors will remain. The assumptions required for averaging to work are just:

$$\begin{gathered} \lim_{n\to\infty} n^{-1}\Lambda'\Lambda = D_\Lambda, \text{where } D_\Lambda \text{ is } r \times r \text{ and has full rank, and} \\ maxeval(\Sigma_e) \leq c < \infty \; \forall \; n, \end{gathered} \tag{C.4}$$

where *maxeval* denotes the maximum eigenvalue; and $\Sigma_e$ is the covariance matrix of $e_t$.

Consider a weighting matrix $W$ (with $W'W/n = I$) such that the factors are estimated as $\hat{F}_t = n^{-1}W'X_t$. If $\lim_{n\to\infty} n^{-1}W'\Lambda = H_{r\times r}$; $H$ has full rank; and conditions in eq. (C.4) are satisfied, then:

$$\hat{F}_t = n^{-1}W'\Lambda F_t + n^{-1}W'e_t \xrightarrow{p} HF_t \text{ as } n \to \infty,$$

where it was used that $n^{-1}W'e_t \xrightarrow{p} 0$ by the weak law of large numbers. Note that without imposing some additional restrictions $F_t$ and $\Lambda_t$ are not identified because matrix $H$ is unknown. Since $H$ is $r \times r$ we need $r^2$ restrictions to identify the factors and their loadings. The usual normalization assumption $n^{-1}\Lambda'\Lambda = I_r$ provides $r(r+1)/2$ restrictions. The remaining $r(r-1)/2$ restrictions are obtained imposing $F'F$ to be diagonal, where $F = (F_1', ..., F_T')$.

The matrix $W$ is not unique and can be selected in many different ways. In the *Principal Components* approach, $W$ is the matrix of eigenvectors of the sample covariance matrix of $X_t$. Specifically, for a given number of factors $k$ (not necessarily equal to $r$) the principal components method estimates the factors and loadings by solving the optimization problem:

$$\underset{F_1,...,F_T,\Lambda}{Min} S_k, \text{ with, } S_k = (nT)^{-1}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t), \tag{C.5}$$

subject to the normalization $n^{-1}\Lambda'\Lambda = I_r$ and the restricting $F'F$ to be diagonal (what is automatically satisfied). The problem can be solved by concentrating out $F_t$. This gives the least squares estimator of $F_t$ given $\Lambda$, so that $\hat{F}(\Lambda) = (\Lambda'\Lambda)^{-1}\Lambda'X_t$. Then, eq. (C.5) can be

rewritten as:

$$\underset{\Lambda}{Min}\ T^{-1}\sum_{t=1}^{T} X_t'[I-(\Lambda'\Lambda)^{-1}\Lambda']X_t$$

But this new problem is equivalent to:

$$\underset{\Lambda}{Max}\ tr\ \{(\Lambda'\Lambda)^{-1/2'}\Lambda'(T^{-1}\sum_{t=1}^{T} X_tX_t')\Lambda(\Lambda'\Lambda)^{-1/2}\}, \tag{C.6}$$

which is also equivalent to $\underset{\Lambda}{max}\Lambda'\sum_{t=1}^{T} X_tX_t'\Lambda$, subject to $n^{-1}\Lambda'\Lambda = I_k$. This final problem is the starting point of principal components analysis, which solution is to set $\hat{\Lambda}$ equal to the eigenvectors of $\hat{\Sigma}_X = X_t'X_t$ corresponding to its $k$ largest eigenvalues. Next, as $\hat{\Lambda}'\hat{\Lambda} = nI_r$, we get $\hat{F}_t = n^{-1}\hat{\Lambda}'X_t$, which are the scaled $r$ principal components.

An simpler way to see the connection between DFM and principal components is the one in Jolliffe (2005). Let $z = A'X$ be the matrix containing all the $n$ principal components of $X$, where $X_{n\times T} = (X_1, ..., X_T)$. Since matrix $A$ contains the eigenvectors of $\hat{\Sigma}_X$, then $X = Az$ due to the orthogonality of $A$. Partitioning $A$ (and $z$) into the first $r$ and last $n-r$ columns (rows):

$$\begin{aligned} X &= (A_r|A^*_{n-r})\begin{pmatrix} z_r \\ \hline z^*_{n-r} \end{pmatrix}, \\ &= A_rz_r + A^*_{n-r}z^*_{n-r} \\ &= \Lambda F + e \end{aligned} \tag{C.7}$$

where $\Lambda = A_r$, $F = z_r$ and $e = z^*_{n-r}A^*_{n-r}$.

Bai and Ng (2008) summarize the properties of the estimated factors and loadings. Briefly, as proved by Bai and Ng (2002) both estimators are consistent (the average squared deviation between the k estimated factors and the space spanned by k of the true factors vanish at rate $min[N,T]$), and they converge to normal distributions. Moreover, for each $t$, estimated factors are $\sqrt{N}$ consistent for the true factor space while for each $i$, estimated factor loadings are $\sqrt{T}$ consistent for the space spanned by the true factor loadings (see the details in chapter 3 of Bai and Ng (2008)).

Finally, given that the covariance matrix of $e_t$ is not assumed to be diagonal, generalized principal components methods have been proposed to take this feature into account. Several approaches have been proposed to make this procedure feasible (see Forni et al. (2005); and Stock et al. (2005)). Nonetheless, empirical applications to real and simulated data do not show

the generalized method to produce better forecasting results systematically (see e.g Boivin and Ng (2005); D'Agostino and Giannone (2012); or Forni et al. (2005)).

#### C.1.1.3 A combination of previous two approaches

This class of methods try to exploit the efficiency of state space models and the ability of principal components to deal with large $n$. For the static representation of the model, in a first step the factors are estimated by principal components. Next, these factors are used to estimate the parameters of the state space representation eq. (C.2) regressing $X_t$ on $\hat{F}_t$, and estimating the VAR for $\hat{F}_t$. Residuals of this regression are used to estimate eq. (C.3).

These parameters 'complete' the state-space model so that improved estimates of the factors can be obtained using the Kalman smoother.

Other possibility is to use the estimated coefficients as starting values for ML estimation of the parameters via the Kalman filter.

### C.1.2 Estimating the number of factors

Bai and Ng (2002) highlight two possible information criteria for determining the number of factors:

$$\begin{aligned} PCP(k) &= S_k + k\overline{\sigma}^2 g(n,T) \\ IC_k &= ln(S_k) + kg(n,T) \end{aligned},$$

where $g(n,T)$ is a penalty function, $S_k$ is given by eq. (C.5) and $\overline{\sigma}^2 = S_{kmax}$ for a certain value of $kmax$. In both cases $\hat{k}$ is determined by minimizing the information criteria over $k$. The authors show that when $g(n,T) \to 0$ and $[min(n,T)]g(n,T) \to \infty$ as $(n,T) \to \infty$ the probability of selecting the correct number of factors tends to one for both criteria.

Bai and Ng (2002) authors consider four possible penalty functions:

$$g_1(n,T) = \frac{n+T}{nT} ln[min(n,T)]$$

$$g_2(n,T) = \frac{n+T}{nT} ln[\frac{nT}{n+T}]$$

$$g_3(n,T) = \frac{ln[min(n,T)]}{min(n,T)}$$

$$g_4(n,T) = (n+T-k)ln(nT)/nT,$$

and state that $g_2(n,T)$ tends to be more stable in empirical applications. Additionally they argue that $g_4(n,T)$ has good properties especially when the errors are cross correlated, but strictly speaking, it does not tends to zero when $(n,T)$ go to infinity if $T = exp(n)$ or $n = exp(T)$. But, as they argue, these configurations of $n$ and $T$ do not seem empirically relevant. Thus, $g_4(N,T)$ is also to be used in practice.

Amengual and Watson (2007) show that both, the $PCP$ and the $IC$ criteria applied to $\hat{S}_k$ also produce consistent estimates of the number of factors.

Another possibility for determining the number of factors is the suggested by Bai and Ng (2007). Based on the observation that residuals' covariance matrix of the VAR in second line of eq. (C.2) has rank q, the authors propose estimating the number of factors analyzing the eigenvalues of the covariance matrix of those residuals. Since this matrix has r-eigenvalues, the smallest $r - q$ are zero. Calling $c_1, .., c_r$ the ordered eigenvalues, the authors define:

$$D_{1k} = (\frac{c_{k_1}^2}{\sum_{j=1}^{r} c_j^2})^{1/2} \quad and \quad D_{1k} = (\frac{\sum_{j=k+1}^{r} c_j^2}{\sum_{j=1}^{r} c_j^2})^{1/2},$$

so that, for $k \geq q$ $D_{1k} = D_{2k} = 0$.

Define now: $M_{nT}(\delta) = \frac{m}{min[n^{1/2-\delta}, T^{1/2-\delta}]}$, $\hat{q}_1$ to be the smallest $k$ such that $\hat{D}_{1k} < M_{nT}$ and $\hat{q}_2$ to be the smallest $k$ such that $\hat{D}_{2k} < M_{nT}$ for some $0 < m < \infty$ and $0 < \delta < 1/2$. Bai and Ng (2007) show that $P(\hat{q}_i = q) \to 1$ as $n, T \to \infty$.

From Monte Carlo experiments, Stock and Watson (2011) conclude that Bai and Ng's (2007) approach has somewhat better finite sample properties than the Amengual and Watson (2007) procedure.

### C.1.3 DFMs with non stationary data

So far we have considered that the data is stationary, so that economic variables would need to be differenced to fit into this framework. In this section we briefly review available methods to estimate non-stationary factors.

Given the model specified in equations eq. (C.2) and eq. (C.3), there could be two sources of non-stationarity in the data. First, polynomial matrix $A(L)$ can contain unit roots, second, polynomials $d_i(L)$ can contain unit roots. In the first case the common factors are non-stationary, whereas in the second the idiosyncrasies are non-stationary.

When estimation is carried out via the Kalman filter, non-stationarity is not a problematic

issue since the filter procedures do not require the data to be stationary.[1] However, in the principal components approach matrix $\hat{\Sigma}_X$ will contain large eigenvalues even if there are no common factors in the data (spurious regression problem).

Peña and Poncela (2006b) propose a procedure for building DFMs when the factors can be non-stationary, but the idiosyncrasies are known to be I(0) and serially uncorrelated. The authors show that when the model has $r$ common factors, $r_1$ of which are non-stationary, the generalized covariance matrix in equation eq. (C.8) converges weakly to a random matrix $\Gamma_x$ as $T$ goes to infinity.

$$C_X(k) = \frac{1}{T^{2d+d'}} \sum_{t=k+1}^{T} (X_{t-k} - \overline{X})(X_t - \overline{X})', \tag{C.8}$$

where $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ and $d'$ can be 1 or 0. Matrix $\Gamma_x$ has $r_1$ eigenvalues larger than zero and $n - r_1$ equal to zero. Therefore, the number of non-stationary common factors can be found as the number of non-zero eigenvalues of $C_X(k)$, and the eigenvectors associated with these eigenvalues are a basis for the space spanned by the factor loading sub-matrix corresponding to the non-stationary factors.

Peña and Poncela (2006b) also propose a $\chi^2$ test statistic for determining the number of common factors (some of which can be non-stationary). Similarly to cointegration and common cycles rank tests, Peña and Poncela propose to compute the squared canonical correlations $(\hat{\lambda}_i)$ between $X_{t-k}$ and $X_t$ and use the following test statistic:

$$Q_{n-r} = -(T-k)\sum_{j=1}^{n-r} log(1-\hat{\lambda}_i) \sim \chi^2_{(n-r)^2} \tag{C.9}$$

When the model for the factors is unknown, the tests need to be performed for different choices of $k$. In a Monte Carlo study the authors show that the procedure is very powerful and has correct size for detecting the number of factors.

Since Peña and Poncela (2006b) work in a framework of small $n$ and large $T$ the principal components procedure is not consistent, so they propose an estimation procedure based on the Kalman filter.

Given the previous results, the authors suggest the following estimation procedure: first determine $\hat{r}$ using $Q_{n-r}$ in eq. (C.9), then consider the first $\hat{r}$ eigenvectors of $C_X(k)$ as an initial

[1] As highlighted by Poncela and Ruiz (2012), when both, the factors and the idiosyncrasies, are not stationary and the vector of idiosyncrasies has at most $n-1$ unit roots the Kalman filter reaches the steady state, meaning that the system matrices converge to constant matrices (see also Harvey, 1990).

estimate of the factor loading matrix ($\tilde{\Lambda}$). Third, use these estimators as initial conditions for the Kalman filter.

In this case that $F_t$ are integrated and $e_{it}$ are stationary, Bai (2004) shows that the estimating common factors as described in §C.1.1.2 delivers faster convergence to the true than in the case of $I(0)$ factors. So the principal components approach can also be used.

## C.2 Appendix to section IV.5.2: Simulation results when the DGP is a DFM

Table C.1: Gauge and Potency of the pairwise approach. DGP: 'Non-Stat-DFM'

| | **Pairwise** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 1.0 | 85.9 | 9.1 | 31.7 | 2.7 | 20.3 | 5.1 | 12.8 | 3.6 | 49.3 |
| Sce 2 | 0.9 | 85.4 | 15.9 | 29.0 | 6.9 | 25.9 | 5.5 | 12.5 | 8.4 | 42.5 |
| Sce 3 | 0.9 | 82.2 | 24.6 | 28.2 | 10.8 | 29.5 | 5.0 | 15.6 | 13.8 | 39.3 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 1.0 | 78.1 | 7.7 | 22.0 | 3.3 | 28.4 | 4.6 | 10.5 | 2.8 | 31.2 |
| Sce 2 | 0.9 | 79.1 | 11.6 | 20.8 | 8.2 | 38.2 | 4.5 | 10.6 | 6.0 | 26.5 |
| Sce 3 | 0.9 | 75.1 | 17.4 | 19.4 | 12.9 | 39.4 | 4.6 | 10.8 | 9.4 | 24.3 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 1.1 | 76.9 | 5.7 | 15.5 | 3.9 | 42.3 | 3.8 | 6.3 | 2.0 | 19.0 |
| Sce 2 | 0.9 | 81.6 | 8.7 | 15.8 | 9.8 | 53.3 | 4.1 | 6.7 | 4.2 | 18.5 |
| Sce 3 | 0.7 | 78.6 | 12.9 | 15.1 | 15.2 | 54.0 | 4.6 | 6.1 | 6.6 | 18.1 |

- Number of experiments: 500.
- For scenarios 1 to 3, $SC$ is 10, 25 and 40 respectively.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- $Z_2$ = number of wrong series included in the corresponding subset.
- $Z_1$ = number of correct series included in the corresponding subset.
- $Nexp$ = number of experiments
- *CT* column represents the dark red box of figure IV.5.
- *CC* column represents the dark blue box of figure IV.5.
- *OnlyCT* column represents box 1 of figure IV.5.
- *OnlyCC* column represents box 8 of figure IV.5.
- *Intersect* column represents box 5 of figure IV.5.

Table C.2: Gauge and Potency of the DFM approach. Two factors. DGP: 'Non-Stat-DFM'

| | **DFM counterpart** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **T=400** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 17.6 | 97.1 | 60.9 | 88.4 | 21.9 | 94.3 | 62.0 | 95.6 | 16.3 | 81.2 |
| Sce 2 | 8.3 | 88.5 | 28.3 | 84.4 | 21.6 | 78.0 | 35.0 | 96.0 | 17.5 | 73.0 |
| Sce 3 | 9.0 | 76.5 | 28.0 | 76.3 | 29.8 | 61.0 | 36.7 | 90.0 | 21.6 | 60.0 |
| | **T=200** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 28.8 | 95.6 | 65.0 | 81.3 | 32.5 | 91.6 | 65.5 | 88.5 | 22.5 | 73.9 |
| Sce 2 | 9.0 | 87.8 | 38.9 | 82.9 | 22.1 | 77.0 | 43.9 | 93.3 | 17.3 | 72.1 |
| Sce 3 | 9.4 | 75.9 | 32.6 | 75.1 | 29.9 | 60.4 | 40.1 | 87.9 | 21.4 | 59.3 |
| | **T=100** | | | | | | | | | |
| | CT | | CC | | OnlyCT | | OnlyCC | | Intersect | |
| | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot | Gauge | Pot |
| Sce 1 | 39.8 | 93.1 | 64.9 | 72.9 | 42.9 | 88.2 | 65.0 | 78.5 | 28.3 | 65.9 |
| Sce 2 | 12.4 | 86.8 | 48.2 | 79.1 | 24.9 | 75.9 | 51.5 | 88.8 | 17.6 | 68.8 |
| Sce 3 | 10.1 | 75.9 | 39.1 | 73.3 | 30.2 | 61.4 | 45.0 | 84.0 | 20.7 | 58.6 |

See notes to table C.1.

*This page intentionally left blank.*

# Chapter V

# Some issues for empirical applications of the pairwise approach with common cycles

## V.1 Introduction

In this short chapter we consider three relevant issues for empirical applications of the pairwise approach with common cycles: the inclusion of a strongly exogenous variable, the extraction of the common cycle inside the 'single-cycle' subsets, and the use of the *parsimonious long lag* structures (PLL) described in §III.4.1.

In §V.2 we analyze how the pairwise approach applied to common cycles has to be modified when an exogenous variable is added to the system. In lemma 1 we show that, unless the strongly exogenous variable is a white noise, one of the *common cycles* in the system is generated only by the short run structure of that series. In a three dimensional sub-system (two components and an exogenous variable, $EXO$), there can be zero, one, two, or three common cycles, i.e., $s = 3$, $s = 2$, $s = 1$, or $s = 0$, respectively. The cases of interests are $s = 2$ and $s = 1$. In §V.2.2 we show that if a component belongs to a sub-system with $s = 1$, then it cannot belong to other with $s = 2$. Based on this result, we design an empirical strategy for dealing with exogenous variables.

After the pairwise procedure is applied and the 'single-cycle' subsets constructed, the common cycle inside each subset has to be estimated to be used in the forecasting equations. In §V.3

we propose two alternatives for doing this.

Finally, in §V.4 we argue that the $PLL$ structures considered in §III.4.1 should not be used when testing for common cycles.

## V.2 Inclusion of an exogenous variable in the sub-systems in which common cycles tests are performed

In this section we analyze how the pairwise approach applied to common cycles has to be modified when an exogenous variable ($EXO$) is added to the system.

### V.2.1 Common cycles and Granger Causality

This sub-section is devoted to show that the following lemma holds.

**Lemma 1** *In a model with WF, if there is a variable which is not Granger caused by any other variable in the system, there are only two possible situations:*

*(a) one of the common cycles is generated exclusively by the not Granger caused variable; or*

*(b) the not Granger caused variable does not have any of the common cycles in the system, hence, its short run dynamics is dramatically limited, meaning that it must be white noise, except for the cointegration relationships.*

**Proof** *Consider the usual VEqM with only one lag (the assumption of only one lag is for notation simplicity and does not entail any loss of generality):*

$$\Delta X_t = c + \alpha\beta' X_{t-1} + \Phi\Delta X_{t-1} + \epsilon_t.$$

*If $\Phi = \delta_\perp \Psi'$, with $\Psi$ having dimensions $n \times (n-s)$, $\delta_\perp$ $n \times (n-s)$, and $0 < s < n$; then, there are $s$ WFs and $(n-s)$ common cycles defined as:*

$$CC_{t-1} = \Psi' \Delta X_{t-1}.$$

*If the $n^{th}$ series in $X_t$ is not Granger caused by any other series, then:*

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1n} \\ \vdots & & & \\ \phi_{n-1,1} & \phi_{n-1,2} & \dots & \phi_{n-1,n} \\ 0 & 0 & \dots & \phi_{n,n} \end{bmatrix} \tag{V.1}$$

*Let $S_\delta$ be a selection matrix that selects the last row and all the columns of $\delta_\perp$. Likewise, let $S_\psi$ be the selection matrix that keeps all rows and the first $(n-1)$ columns of $\Psi'$. Define now $\delta^*_{\perp_{1\times(n-s)}} = S_\delta \delta_\perp$, and $\Psi^{*'}_{(n-s)\times(n-1)} = S_\psi \Psi'$. For the last row of $\Phi$ being the one in eq. (V.1) we need:*

$$\delta^*_{\perp_{1\times(n-s)}} \Psi^{*'}_{(n-s)\times(n-1)} = 0_{1\times(n-1)} \tag{V.2}$$

*There are three possible solutions for eq. (V.2):*

*(i) One possible solution is $\delta^*_\perp = 0_{1\times(n-s)}$. This solution implies that the only short run movements of $X_{nt}$ are those caused by equilibrium adjustments. Note that if $X_{nt}$ is weakly exogenous to $\beta$, this solution implies that this series is white noise.*

*(ii) Another possibility could be $\Psi^{*'} = 0_{(n-s)\times(n-1)}$. This solution implies that $rank(\Psi) \leq 1$, so it would be possible only when $n - s = 1$, for $\Psi$ must have full column rank. Since the only non-zero column of $\Psi'$ is the last one, the common cycles is generated exclusively by the $n^{th}$ series.*

*(iii) $\delta^*_\perp \neq 0$; $\Psi^{*'}_\perp \neq 0$; and $\delta^*_\perp \Psi^{*'} = 0_{1\times(n-1)}$. This is the most general solution. When $(n-s) = 1$, this solution is not possible because, with $\delta^*_\perp \neq 0$, the only possibility to satisfy eq. (V.2) is $\Psi^{*'} = 0$ (see below).*

*Solution (i) proves item b, and solution (ii) proves item a. We now show that under solution (iii), case item a must happen. For doing this, we show that the last row of $\Psi^{*'}$ have to be full of zeros so that one of the common cycles is generated only by the last series.*

*We consider the two possibilities of interest for the pairwise approach: $s = 1$, or $1 < s < n$.*

#### V.2.1.1 First case: $s = 1$

*In this case $\Psi^{*'}$ has dimensions $(n-1) \times (n-1)$. Since solution (iii) requires $\Psi^{*'}$ to have reduced rank, it implies $|\Psi^{*'}| = 0$.*

*It can be shown that* $\Psi'_{(n-s)\times n}$ *can always be normalized as* $\tilde{\Psi}'_{(n-s)\times n} = H_{(n-s)\times(n-s)}\Psi' = [\bar{\Psi}'_{(n-s)\times s}, I_{(n-s)}]$.

*Since* $s = 1$, *we have:*

$$\tilde{\Psi}' = \begin{bmatrix} \tilde{\psi}'_{11} & 1 & 0 & 0 & \dots & 0 \\ \tilde{\psi}'_{21} & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ \tilde{\psi}'_{(n-1),1} & 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{(n-1)\times n} .$$

*The corresponding* $\tilde{\Psi}^{*'}$ *is therefore:*

$$\tilde{\Psi}^{*'} = \begin{bmatrix} \tilde{\psi}'_{11} & 1 & 0 & 0 & \dots & 0 \\ \tilde{\psi}'_{21} & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ \tilde{\psi}'_{(n-2),1} & 0 & 0 & 0 & \dots & 1 \\ \tilde{\psi}'_{(n-1),1} & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{(n-1)\times(n-1)} .$$

*Condition (iii) requires the determinant of the* $\tilde{\Psi}^{*'}$ *to be equal to zero. Note now a that necessary and sufficient condition for* $\tilde{\Psi}^{*'}$ *to have reduced rank is* $\tilde{\psi}'_{(n-1),1} = 0$. *Hence, for having* $|\tilde{\Psi}^{*'}| = 0$ *we need the last row of* $\tilde{\Psi}^{*'}$ *to be full of zeros — or, equivalently, the last row of* $\tilde{\Psi}'$ *to be full of zeros except for the* $n^{th}$ *entry —, what implies that one of the* $n-1$ *common cycles is generated exclusively by the* $n^{th}$ *series.*

#### V.2.1.2 Second case: $1 < s < n$

*In this case* $\Psi^{*'}$ *is not square, and the condition* $|\Psi^{*'}| = 0$ *makes no sense. Again, we can normalize* $\Psi'_{(n-s)\times n}$ *as* $\tilde{\Psi}'_{(n-s)\times n} = [\bar{\Psi}'_{(n-s)\times s}, I_{(n-s)}]$. *Using this normalization condition (iii) becomes:*

$$\tilde{\delta}^*_{\perp} \neq 0,\ \tilde{\delta}^*_{\perp}\tilde{\Psi}^{*'} = 0_{1\times(n-1)}, \tag{V.3}$$

*where* $\tilde{\delta}^*_{\perp}$ *is the vector of coefficients that corresponds to the new normalization, and:*

$$\tilde{\Psi}^{*'} = \begin{bmatrix} \tilde{\psi}'_{11} & \dots & \tilde{\psi}'_{1s} & 1 & 0 & 0 & \dots & 0 \\ \tilde{\psi}'_{21} & \dots & \tilde{\psi}^{*'}_{2s} & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & \\ \tilde{\psi}'_{(n-s-1),1} & \dots & \tilde{\psi}'_{(n-1),s} & 0 & 0 & 0 & \dots & 1 \\ \tilde{\psi}'_{(n-s),1} & \dots & \tilde{\psi}'_{(n-s),s} & 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{(n-s)\times(n-1)} . \tag{V.4}$$

*Define now* $\tilde{\Psi}^*_+ = \tilde{\Psi}^*(\tilde{\Psi}^{*'}\tilde{\Psi}^*)^{-1}$. *Since* $\tilde{\Psi}^{*'}\tilde{\Psi}^*_+ = I_{(n-s)\times(n-s)}$, *condition eq. (V.3) requires* $|\tilde{\Psi}^{*'}\tilde{\Psi}^*| = 0$, *what implies that* $\tilde{\Psi}^*_+$ *does not exist.*

*For* $|\tilde{\Psi}^{*'}\tilde{\Psi}^*| = 0$ *we need* $\tilde{\Psi}^{*'}\tilde{\Psi}^*$ *to have reduced rank. Using that* $\tilde{\Psi}^{*'}$ *has dimension* $(n-s)\times(n-1)$ *and that* $(n-s) < (n-1)$, *it follows that if* $\tilde{\Psi}^{*'}$ *has full row rank, then,* $\tilde{\Psi}^{*'}\tilde{\Psi}^*$ *will have full rank. Therefore, we need* $\tilde{\Psi}^{*'}$ *not to have full row rank for condition eq. (V.3) to be satisfied.*

*Note now that from eq. (V.4), the only possibility for rank*$(\tilde{\Psi}^{*'})$ *to be smaller than* $n-s$ *is that* $\tilde{\psi}'_{(n-s),1} = ... = \tilde{\psi}'_{(n-s),s} = 0$, *and the proof is complete* ■

#### V.2.1.3 Two particular cases of interest for the pairwise procedure

**$n = 3$, $s = 2$ $(n - s = 1)$**

The three dimensional model is:

$$\Delta X_t = c + \alpha\beta' X_{t-1} + \delta_\perp \Psi' \Delta X_{t-1} + \epsilon_t \tag{V.5}$$

where, $\delta_\perp = [\delta_{\perp_1}, \delta_{\perp_2}, \delta_{\perp_3}]'$ and $\Psi' = [\psi_1, \psi_2, \psi_3]$.

If $X_{3t}$ is not Granger caused by any of the two other variables we need:

$$\delta_\perp \Psi' = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{2,1} & \phi_{2,2} & \phi_{2,3} \\ 0 & 0 & \phi_{3,3} \end{bmatrix} \tag{V.6}$$

The sub-matrices $\delta^*_{\perp_{1\times(n-s)}}$ and $\Psi^{*'}_{(n-s)\times(n-1)}$ defined above are in this case:

$$\delta^*_\perp = \delta_{\perp_3}, \text{ and } \Psi^{*'} = [\psi_1, \psi_2].$$

The three possible solutions analyzed above are:

(i) $\delta_{\perp_3} = 0$. But this implies that $\Delta X_{3t}$ is white noise (if it is also weakly exogenous to $\beta$), or that its short run movements are completely explained by equilibrium corrections.

(ii) $\psi_1 = \psi_2 = 0$. This implies that the unique common cycle in the system is generated by $X_{3t}$, and:

$$\delta_\perp \Psi' = \begin{bmatrix} 0 & 0 & \phi_{13} \\ 0 & 0 & \phi_{2,3} \\ 0 & 0 & \phi_{3,3} \end{bmatrix}$$

(iii) Since $n-s = 1$, requiring the last row of $\Psi^{*'}$ being full of zeros is equivalent to the previous condition.

**$n = 3$, $s = 1$ $(n - s = 2)$**

In this case we have

$$\delta_\perp = \begin{bmatrix} \delta_{\perp_{11}} & \delta_{\perp_{12}} \\ \delta_{\perp_{21}} & \delta_{\perp_{22}} \\ \delta_{\perp_{31}} & \delta_{\perp_{32}} \end{bmatrix}, \textit{and } \Psi' = \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \end{bmatrix} \tag{V.7}$$

Again, if $X_{3t}$ is not Granger caused by $X_{1t}$ or $X_{2t}$, eq. (V.6) holds.

The sub-matrices $\delta^*_{\perp_{1\times(n-s)}}$ and $\Psi^{*'}_{(n-s)\times(n-1)}$ defined above are in this case:

$$\delta^*_\perp = [\delta_{\perp_{31}}, \delta_{\perp_{32}}], \textit{ and } \Psi^{*'} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}.$$

Once again, solution (i) strongly restricts the short run dynamics of $X_{3t}$, and solution $(ii)$ is not possible for $\Psi$ must have full column rank.

For solution (iii) we need $\delta^*_\perp \neq [0,0]'$ and $\delta^*_\perp \Psi^{*'} = [0,0]$. Thus, we need $|\Psi^{*'}| = 0$, which implies $\psi_{11}\psi_{22} = \psi_{21}\psi_{12}$. As aforementioned, $\Psi'$ can always be normalized as:

$$\tilde{\Psi}' = \begin{bmatrix} \tilde{\psi}_{11} & 1 & 0 \\ \tilde{\psi}_{21} & 0 & 1 \end{bmatrix}, \tag{V.8}$$

so that:

$$\tilde{\Psi}^{*'} = \begin{bmatrix} \tilde{\psi}_{11} & 1 \\ \tilde{\psi}_{21} & 0 \end{bmatrix},$$

and the condition for $|\tilde{\Psi}^{*\prime}| = 0$ is $\tilde{\psi}_{21} = 0$, i.e, the last row of $\tilde{\Psi}^{*\prime}$ begin full of zeros. Therefore, one of the two cycles is generated exclusively by short run dynamics of $X_{3t}$ and the other by a linear combination of the short run dynamics of $X_{1t}$ and $X_{2t}$.

### V.2.2 'Mixed' sub-systems

When including $EXO$ in the sub-systems there are four possible outcomes of the common cycles tests:

a. There are no common cycles restrictions ($s = 0$).

b. There is only one common cycle restriction ($s = 1$).

c. There are only two common cycles restrictions ($s = 2$).

d. There are three common cycles restrictions (after correcting by their cointegration relationships all the variables are white noise, $s = 3$).

Cases (a) and (d) are not considered in the construction of the 'single-cycle' subsets so we focus on cases (b) and (c). Restricting our attention to those cases, the general situation is that there will be some systems with $s = 1$ and some others with $s = 2$. If $EXO$ is strongly exogenous and it is not white noise, in the first case there are two common cycles, one determined by the $EXO$ and the other by a combination of the two components (see lemma 1). When $s = 2$ there is only one common cycle generated by $EXO$.

We now show that, for one component, finding $s = 1$ and $s = 2$ in two different sub-systems is incompatible. The only assumptions required for this result to hold are that $EXO$ is strongly exogenous and that it is not white noise.

Consider the case that for the sub-system $(X_{1t}, X_{2t}, EXO_t)$ we find $s = 2$ (call this as *sub-system c*) but for sub-system $(X_{1t}, X_{3t}, EXO_t)$ we find $s = 1$ (*sub-system b*).

From *sub-system c* we can write:

$$\begin{aligned} \Delta X_{1t} &= \delta^c_{\perp_{11}} CC_{exo,t} + \boldsymbol{\alpha_1}\beta' X_{t-1} + \epsilon^c_{1t} \\ \Delta X_{2t} &= \delta^c_{\perp_{21}} CC_{exo,t} + \boldsymbol{\alpha_2}\beta' X_{t-1} + \epsilon^c_{2t} \end{aligned},$$

where $\boldsymbol{\alpha_i}$ stands for the $i^{th}$ row of matrix $\alpha$, and $X_t$ is the $n \times 1$ vector of all the components (recall from §IV.4.1 that we include all cointegration relationships in the WF tests).

From *sub-system b*:

$$\Delta X_{1t} = \delta^b_{\perp_{11}} CC_{x_{13},t} + \delta^b_{\perp_{12}} CC_{exo,t} + \boldsymbol{\alpha_1}\beta' X_{t-1} + \epsilon^b_{1t}$$
$$\Delta X_{3t} = \delta^b_{\perp_{21}} CC_{x_{13},t} + \delta^b_{\perp_{22}} CC_{exo,t} + \boldsymbol{\alpha_3}\beta' X_{t-1} + \epsilon^b_{3t}$$

Equalizing the first equations of the two sub-systems:

$$\delta^c_{\perp_{11}} CC_{exo,t} + \epsilon^c_{1t} = \delta^b_{\perp_{11}} CC_{x_{13},t} + \delta^b_{\perp_{12}} CC_{exo,t} + \epsilon^b_{1t},$$

from where:

$$(\delta^c_{\perp_{11}} - \delta^b_{\perp_{12}}) CC_{exo,t} - \delta^b_{\perp_{11}} CC_{x_{13},t} = \epsilon^b_{1t} - \epsilon^c_{1t}. \tag{V.9}$$

Without any restriction in the coefficients, eq. (V.9) is a contradiction because it implies that there exist a linear combination of two independent cycles which is white noise.

For eq. (V.9) to be *possible* two conditions are required:

(i) $\delta^c_{\perp_{11}} = \delta^b_{\perp_{12}}$, *and*

(ii) $\delta^b_{\perp_{11}} = 0$

The two conditions are required at the same time for with only one of them one of the common cycles would be a white noise.

Condition (ii) implies that the coefficient of $\Delta X_{1t}$ on the cycle generated by $EXO$ is equal to 0. Thus, if we require the $\delta' s$ associated to the components to be different from zero, sub-systems $b$ and $c$ cannot happen at the same time.

This result implies that when $EXO$ is included (and it is strongly exogenous and is not white noise) the search of 'single-cycle' subsets should not mix sub-systems with different number of WF relationships.

When sub-systems have two WF restrictions (one common cycle) we are in the 'regular' case and 'single-cycle' subsets can be formed as usual.

We now consider the case where the sub-systems have only one WF relationship.

### V.2.3 Sub-systems with s=1

Consider the following two sub-systems:

$$\begin{aligned}\Delta X_{1t} &= \delta^1_{\perp_{11}} CC_{x_{12},t} + \delta^1_{\perp_{12}} CC_{exo,t} + \boldsymbol{\alpha_1}\beta' X_{t-1} + \epsilon^1_{1t} \\ \Delta X_{2t} &= \delta^1_{\perp_{21}} CC_{x_{12},t} + \delta^1_{\perp_{22}} CC_{exo,t} + \boldsymbol{\alpha_2}\beta' X_{t-1} + +\epsilon^1_{2t}\end{aligned} \tag{V.10}$$

$$\begin{aligned}\Delta X_{1t} &= \delta^2_{\perp_{11}} CC_{x_{13},t} + \delta^2_{\perp_{12}} CC_{exo,t} + \boldsymbol{\alpha_1}\beta' X_{t-1} + \epsilon^2_{1t} \\ \Delta X_{3t} &= \delta^2_{\perp_{21}} CC_{x_{13},t} + \delta^2_{\perp_{22}} CC_{exo,t} + \boldsymbol{\alpha_3}\beta' X_{t-1} + \epsilon^2_{3t}\end{aligned} \tag{V.11}$$

From the first line in eq. (V.11):

$$CC_{x_{13},t} = \frac{\Delta X_{1t} - \delta^2_{\perp_{12}} CC_{exo,t} - \boldsymbol{\alpha_1}\beta' X_{t-1} - \epsilon^2_{1t}}{\delta^2_{\perp_{11}}} \tag{V.12}$$

Plugging eq. (V.12) in the second line of eq. (V.11):

$$\Delta X_{3t} = \frac{\delta^2_{\perp_{21}}}{\delta^2_{\perp_{11}}}(\Delta X_{1t} - \delta^2_{\perp_{12}} CC_{exo,t} - \boldsymbol{\alpha_1}\beta' X_{t-1} - \epsilon^2_{1t}) + \delta^2_{\perp_{22}} CC_{exo,t} + \boldsymbol{\alpha_3}\beta' X_{t-1} + \epsilon^2_{3t},$$

or in a simpler form:

$$\Delta X_{3t} = \lambda_1 \Delta X_{1t} - \lambda_2 CC_{exo,t} + \boldsymbol{\lambda_3}\beta' X_{t-1} + e_{3t}, \tag{V.13}$$

where $\lambda_1 = \frac{\delta^2_{\perp_{21}}}{\delta^2_{\perp_{11}}}$; $\lambda_2 = \delta^2_{\perp_{22}} - \frac{\delta^2_{\perp_{21}}\delta^2_{\perp_{12}}}{\delta^2_{\perp_{11}}}$; $\boldsymbol{\lambda_3} = \boldsymbol{\alpha_3} - \lambda_1\boldsymbol{\alpha_1}$; and $e_{3t} = \epsilon^2_{3t} - \lambda_1\epsilon^2_{1t}$.

Plugging the first equation of eq. (V.10) into eq. (V.13) we get:

$$\Delta X_{3t} = \lambda_1(\delta^1_{\perp_{11}} CC_{x_{12},t} + \delta^1_{\perp_{12}} CC_{exo,t} + \boldsymbol{\alpha_1}\beta' X_{t-1} + \epsilon^1_{1t}) - \lambda_2 CC_{exo,t} + \boldsymbol{\lambda_3}\beta' X_{t-1} + e_{3t},$$

or in simpler terms;

$$\Delta X_{3t} = \gamma_1 CC_{x_{12},t} + \gamma_2 CC_{exo,t} + \boldsymbol{\gamma_3}\beta' X_{t-1} + e^*_t,$$

where; $\gamma_1 = \frac{\delta^2_{\perp_{21}}\delta^1_{\perp_{11}}}{\delta^2_{\perp_{11}}}$; $\gamma_2 = \frac{\delta^2_{\perp_{11}}\delta^1_{\perp_{12}} - \delta^2_{\perp_{22}}\delta^2_{\perp_{11}} + \delta^2_{\perp_{21}}\delta^2_{\perp_{12}}}{\delta^2_{\perp_{11}}}$, $\boldsymbol{\gamma_3} = \frac{\delta^2_{\perp_{21}}}{\delta^2_{\perp_{11}}}\boldsymbol{\alpha_1} + \boldsymbol{\alpha_3} - \lambda_1\boldsymbol{\alpha_1}$; and $e^*_t = \epsilon^2_{3t} + \frac{\delta^2_{\perp_{21}}}{\delta^2_{\perp_{11}}}(\epsilon^1_{1t} - \epsilon^2_{1t})$.

Therefore, provided that $\delta^2_{\perp_{11}} \neq 0$, $\Delta X_{3t}$ can also be written as a linear combination of the two common trends in sub-system eq. (V.10) plus a white noise component. Thus, 'single-

cycle' subsets can be formed considering sub-systems with only one WF restriction. The only difference with the original case is that in these sets there are two common cycles, one determined exclusively by $EXO$ and the other by the components.

The condition $\delta^2_{\perp_{11}} \neq 0$ means that the $CC_{x_{13},t}$ is present in $\Delta X_{1,t}$. Thus, sub-systems in which the common cycle does not affect both components are to be disregarded.

### V.2.4 Empirical Strategy

Given previous results we can proceed as follows:

1. Focus first on sub-systems with $s = 2$ and look for 'single-cycle' subsets.
2. Disregard series in some subset of the previous step, and look for subsets with $s = 1$. In this case, each subset will have two common cycles, one generated by the exogenous variable, and the other by the two components.

Note that the order of the two steps should not matter given the incompatibility of 'mixed' systems.

## V.3 Extraction of the common cycle for the forecasting equations

Once the 'single-cycle' subsets have been formed, the common cycle has to be extracted to be used in the forecasting models. If the subsets are *small* this is not a problem as the multivariate models can be estimated. In this brief section we propose two possible solutions for not small subsets.

### V.3.1 Solution 1

As inside 'single-cycle' subsets there is only one (two) common cycle(s) and we have $sc(sc-1)/2$ different estimations of this cycle(s), we could extract the first (first two) principal component(s) and use it (them) as the estimator(s) of the common cycle(s).

### V.3.2 Solution 2

Let the multivariate model for the series in the subset to be:

$$\Delta X_t = c + \alpha\beta' X_{t-1} + \delta_\perp \Psi'[\Delta X'_{t-1}, ..., \Delta X'_{t-k}]' + \epsilon_t, \tag{V.14}$$

where $X_t$ has dimensions $sc \times 1$; and $\delta'_{s\times sc}$ contains the $s$ WF restrictions, with $s = sc - 1$ (or $s = sc - 2$ in the case two common cycles). Instead of the cointegration relationships between the series in $SC$ we could have included cointegration relationships with all the components and the argument would keep exactly the same.

Without loss of generality, matrix $\delta'$ can be normalized as $[\tilde{\delta}'_{s\times sc-s}, I_s]$. Then, the $s \times (sc-s)$ unknown coefficients can be obtained from all the bi-variate models that contain the first $sc-s$ series. Thus, by means of the pairwise procedure we get an estimator of $\delta$ and we can construct $\delta_\perp$, to get:

Define $\delta^+_\perp = (\delta'_\perp \delta_\perp)^{-1}\delta'_\perp$, and multiply both sides of eq. (V.14) by $\delta^+_\perp$.

$$\delta^+_\perp \Delta X_t = c^* + \delta^+_\perp \alpha\beta' X_{t-1} + \Psi'[\Delta X'_{t-1}, ..., \Delta X'_{t-k}]' + \epsilon^*_t. \tag{V.15}$$

Since after the pairwise procedure we can construct an estimator of $\delta^+_\perp$, $\hat{\Psi}$ can be obtained from regression eq. (V.15). Noting that $\delta^+_\perp$ has dimension $(sc - s) \times sc$, that $sc - s = 1$ (or 2), and that the cointegration relationships $\beta X_{t-1}$ where previously estimated, eq. (V.15) is a single equation (system of two equations) that can be estimated by (multivariate) OLS and $k$ selected by some information criteria.

For the case of $sc - s = 1$, in eq. (V.15) there are $1 + r + sc \times k$ unknown coefficients, so that this solution will not be useful when $sc$ is large relatively to $T$.

### V.3.3 Simulation results

In order to assess the performance of the two previous solutions we perform a small Monte Carlo study. In each replica we focus on the first 20 series generated in §IV.5 for scenario 3. In that scenario, the first 25 series shared a unique common cycle, so that the 20 series we are keeping share a unique common cycle.

Using those 20 series, we apply the pairwise procedure, construct the *almost* fully connected subset, estimate the common cycle using the aforementioned solutions 1 and 2, and call those estimations $\hat{CC}_{1t}$ and $\hat{CC}_{2t}$. Next, we run the regressions:

$$CC_t = c + \alpha \hat{CC}_{it} + \epsilon_t \quad for \quad i = 1, 2, \tag{V.16}$$

and compute the $R^2$ in both cases. Table V.1 shows, that the second solution is very accurate and clearly dominates the first one.

Table V.1: Comparision of procedures for extracting the common cycle. $R^2$ from regression eq. (V.16)

| | **Sol 1** | **Sol 2** |
|---|---|---|
| | **Mean $R^2$** | |
| $T = 400$ | 0.602 | 0.970 |
| $T = 200$ | 0.607 | 0.959 |
| $T = 100$ | 0.646 | 0.910 |
| | **Median $R^2$** | |
| $T = 400$ | 0.604 | 0.994 |
| $T = 200$ | 0.602 | 0.987 |
| $T = 100$ | 0.638 | 0.964 |

## V.4 Parsimonious Long Lag structures and common cycles tests

In the same way that in §III.4.1 we extended cointegration tests by the consideration of *Parsimonious Long Lag* ($PLL$) structures, we may also be tempted to consider this extension for common cycles' tests. In the following lines we argue that this should not be done.

Assume that the correct model is:

$$\Delta X_t = \Phi_1 \Delta X_{t-1} + \Phi_2 \Delta X_{t-2} + \Phi_3 \Delta X_{t-3} + \epsilon_t.$$

Assume further that $\Phi_1$ and $\Phi_2$ have a common left null space but $\Phi_3$ has full rank. Then, the presence of $\Phi_3$ precludes the existence of common cycles.

Assume now that we estimate the model using a $PLL$ structure, including, for instance, $\Delta_3 X_{t-3}$ instead of $\Delta X_{t-3}$. Given that the $\Phi_4$ to $\Phi_6$ are zero, the estimated coefficient matrix associated $\Delta_3 X_{t-3}$ could be close to zero. Therefore, in the model with $PLL$ we could wrongly find common cycles structures.

This simple example constitutes an argument for not considering $PLL$ structures in common cycles tests.

# Chapter VI

# Empirical applications

## VI.1 Introduction

In this chapter we update the application of chapter III with the consideration of common cycles and include a complete application for the US Industrial Production Index (IPI).

As mentioned in the introduction of this thesis, our main objective is to model and forecast all the components of an aggregate and we do this with a single-equation approach that incorporates long and short run restrictions discovered with the pairwise approach. By means of the pairwise approach we formed what we called 'fully cointegrated' and 'single-cycle' subsets, that have the property that all the series inside them share a single common trend and a single common cycle, respectively. Each series can belong to both, a 'fully cointegrated' and a 'single-cycle' subset. The definition of 'common-cycle' we are considering is the *weak form of serial correlation common feature* (WF) proposed by Hecq et al. (2006). This definition requires the existence of a linear combination that cancels out all the short-run structure of the series except for the cointegration relationships (if they exist), see §IV.2.3 for the technical details.

For building the single-equation models we use the automatic model selection algorithm *Autometrics*. Starting from a General Unrestricted Model (GUM) and using a multiple path search, this algorithm reduces the GUM to a simpler model that encompasses it and passes a battery of diagnostic tests (see Doornik (2009)). Additionally, we consider the Impulse Indicator Saturation (IIS) procedure to control for data irregularities in the models (for a brief description of IIS see §III.2.1, or Santos et al. (2008) for a detailed analysis).

Although we consider more than one possible GUM, a representative example for a series

that belongs to a 'fully cointegrated' subset of size $m$, would be:

$$\Delta x_t = c + \sum_{i=1}^{m-1} \alpha_i CR_{i,t-1} + \sum_{i=1}^{11} \gamma_i S_{it} + \sum_{i=1}^{K} \phi_i \Delta x_{t-i} + \sum_{j=1}^{Q} \sum_{i=1}^{M} \lambda_i R_{j,t-i} + \epsilon_t,$$

where $S_{it}$ represents the $i_{th}$ centered seasonal dummy, $CR_{i,t}$ a cointegration relationship; and $R_{i,t}$ other possible regressors. If $x_t$ also belongs to a 'single-cycle' subset, the model will be:

$$\Delta x_t = c + \sum_{i=1}^{m-1} \alpha_i CR_{i,t-1} + \sum_{i=1}^{11} \gamma_i S_{it} + \psi CC_{t-1} + \epsilon_t,$$

where $CC_t$ represents the common cycle.

Since for each component, the process of building the econometric model is subject to a set of diagnostic tests included *Autometrics*, we can conclude that they are reasonable for empirical applications. Additionally, as the basic components aggregate to the CPI or the IPI, we can apply another test to the models for the disaggregates. It consist of comparing the forecast of the aggregate obtained by aggregating the forecasts of the components with the forecasts from a scalar model for the aggregate forecast. We denote the indirect approach by *I-PW* (*PW* stands for *pairwise*) and the direct one by *D*. The latter is our baseline model.

The pairwise strategy (*I-PW*) would not only provide models to analyze all the components, but it could also be an instrument to obtain more accurate forecasts of the aggregate. This could be so because it incorporates more information than the corresponding direct forecast and could palliate the curse of dimensionality in the number of variables by considering restrictions between them. Therefore, our approach to forecast the aggregate is an intermediate one between the direct approach and the vector-model approach (a full information method, that in our case of interest is not feasible).

Given our interest in forecasting all the components, the comparison with direct approaches should not be used as a definitive criterion for assessing the forecasting performance of our procedure, we should use some disaggregated baseline. Therefore, we also compare the forecasting performance of *I-PW* with the disaggregated forecasts using univariate models for each basic component, denoted as *I*.

The rest of the chapter is organized as follows. In §VI.2 we describe the empirical design of the common cycles tests ( for the empirical design of Johansen's tests see §III.4) and in §VI.3 we describe the forecasting exercises. In §VI.4 we complete the results of chapter III with the

common cycles tests and perform the forecasting exercise. In §VI.5 we perform a complete application for the US IPI including the analysis of outliers, cointegration and common cycles tests, and the forecasting exercise. Finally we devote §VI.6 for the conclusions of the chapter.

## VI.2 Empirical design of the common cycles' tests

For grouping the components by blocks we consider the strategy described in chapter IV augmented by the *relaxation procedure* also described in that chapter. We look for $WF$ restrictions between all possible pairs of components correcting the tests with all the common unit root restrictions that are relevant for the series in the pair. If the pair is cointegrated, we include this relationship. Additionally, if some of them belong to a fully cointegrated subset, we include all the cointegrated relationships of the subsets. Other cointegration relationships outside fully cointegrated subsets are included only if they appear when using a significance level of 0.5%. This procedure ensures transitivity. We disregard blocks with less than four series. In the *relaxation step* we admit consider series that failed to show a common cycle with at most 30% of the series in the set at the original significance level of 10% but showed a common cycle at 0.5%.

As we did for long run restrictions, we disregard estimated short run restrictions that do not fulfill certain quality controls. In this case the conditions are: (i) coefficients of the $\delta_{W\perp}$ are different form zero, and (ii) $\Psi'_W$ is stable over time (recall the model with WF: $\Delta X_t = \alpha\beta' X_{t-1} + \delta_{W\perp}\Psi'_W(\Delta X'_{t-1}, ..., \Delta X'_{t-k+1})' + \epsilon_t$).

The first condition ensures that none of the series is a white noise (after correcting for long run restriction, outliers and seasonality). Then, the estimated common cycle is 'truly' common to both series. Test are performed at 10% of significance.

Regarding the second condition, we evaluate the stability of both, the smallest eigenvalue and the coefficients in $\Psi'_W$ by adapting the procedures in Hansen and Johansen (1999) for the case of common cycles. As we did for cointegration, we assess the eigenvalues by the *fluctuation test*, and compare the coefficients $\psi'$ in the evaluation period with the full sample estimation. In both cases we consider forward recursive tests at 5% of significance based on the concentrated model and the evaluation period is the last five years of the sample.

Apart from the cointegration relationships, in all common cycles' tests we include the relevant impulses to both series; centered seasonal dummies, and, in the case of the IPI, calendar

effects. The lag length is determined in a reduced rank model that includes the cointegration relationships, the impulses, the seasonal dummies and the calendar effects when appropriate.

In the case of 'fully cointegrated' subsets we considered two extensions; the inclusion of an exogenous variable (the Real effective exchange Rate, *REER*), and Parsimonious Long Lag structures (see §III.4 for the details). As we argued in §V.4, PLL structures should not be used when testing for common cycles, so we disregard this extension.

We focus then on the inclusion of the REER. In §V.2 we analyzed how the pairwise approach applied to common cycles has to be modified when an exogenous variable is added to the systems. There, we showed that in a model with *WF*, if there is a variable which is not Granger caused by any other variable in the system, there are only two possible situations: (a) one of the *common cycles* is generated only by the short run structure of the not Granger caused variable; or (b) the not Granger caused variable cannot not have any of the common cycles in the system, hence, its short run dynamics are *dramatically* limited. If in addition to not being Granger caused the variable is weakly exogenous, situation (b) implies that the variable must be white noise.

In a three dimensional sub-system (two prices and the REER), there are two situations of interest; $s = 2$ (one common cycle), or $s = 1$ (two common cycles). Using previous result, assuming that the third variable is not Granger caused by the prices and that it is not white noise, we also showed that if a price belongs to a sub-system with $s = 1$, then it cannot belong to another with $s = 2$.

Hence, in order to proceed as we did in the case of cointegration, we need the *REER* not to be white noise or Granger caused by the components. *REER's* correlogram shows significant values at the first two lags and those lags also retained by *Autometrics* (with p-values close to zero) in an uni-variate model with Impulse Indicator Saturation. Therefore, the *not white noise* assumption seems valid.

Regarding Granger causation, we estimated $N$ models (one for each component) with *log REER's* first difference as dependent variable, and lags one to four of both, the dependent variable and the first difference of the log components. Under independence, if none of the components Granger causes the *REER*, for each lag we would get a proportion $\varphi$ of significant coefficients, with $\varphi$ being the significance level of the tests.

For the CPI, the observed proportions of p-values lower than 5% are 2.4%, 2.4%, 3.6%, and

6.0% for lags one to four, respectively. The test for the null that these proportions are equal to 5% vs being larger is not rejected in all cases. Thus, we also conclude that the $REER$ is not Granger caused by the prices.

For the IPI, the observed proportions of p-values lower than 5% are 3.7% are 7.4%, 5.9%, and 5.9% for lags one to four, respectively. A test for the null that these proportions are equal to 5% vs being larger is not rejected in all cases. Thus, we conclude that the $REER$ is not Granger caused by the sectoral industrial production indexes.

As we did in cointegration tests, we include the $REER$ in a sequential fashion; first, we carry out the search of 'single-cycle' subsets not including the $REER$ and then, using only the series not belonging to any block, we repeat the the procedure but including the $REER$.

## VI.3 Design of the forecasting comparison

With reference to the indirect disaggregated procedures, we initially consider four broad possibilities that differ in the regressors included in the formulation of the initial General Unrestricted Model (GUM). Apart from own lags and seasonal dummies, we may include (a) no other regressor; (b) lags of the aggregate; and (c) lags of the official sub-aggregates corresponding to a breakdown of the aggregate in eight and seven categories for the CPI and IPI, respectively.

For each of these three possibilities, series with no common unit roots or common cycles restrictions can be modeled individually or all together with a scalar model for the sub-aggregate *rest*. Abusing notation, we label this last possibility as *GP*, for Guerrero and Peña (2003).

So, in principle, we have six different possibilities. For each of these alternatives we could consider both, common unit roots and common cycles restrictions, only unit roots, only cycles, or none. Noting that when considering neither common unit roots nor common cycles the option *GP* is not available, we end up with end up with 21 different indirect possibilities.

When not using common unit roots or cycles we add an additional possibility consisting of including dynamic factor models estimated from all the disaggregates (I-DFM), what rises the number of options to 22.

Finally, as argued in chapter II, different normalizations of the unit-root restrictions may lead to different forecasting accuracy (by changing the amount common unit roots restrictions relevant to each component), so we consider three alternative normalizations. In the first one, restrictions are expressed as deviations of all the variables with respect to the dependent variable

in the corresponding equation (a different normalization is used in each equation). In the second normalization, we randomly select a variable and express all restrictions as deviations from the chosen variable. In the third one, restrictions are normalized with respect to the sub-aggregate formed by the series in the corresponding subset.

Table VI.1 includes a summary of the indirect forecasting alternatives (we consider only the case in which no lags of other components or sub-aggregates are included in the model — case (a) above —, the other three cases are simple extensions).

The equations in table VI.1 represent the initial GUMs from where models are selected using *Autometrics* with Impulse Indicator Saturation. We select the impulses in two steps. First we use a target gauge of 0.5% to select variables, lags and impulses, and store the retained impulses. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.

As explained in §III.6.1, the N components we are dealing with may not represent 100% of the aggregate. For the case of the CPI, after aggregating the components' forecasts we get a total weight around 90%, call this sub-aggregate $Agg^*$. To forecast the CPI we consider the following GUM:

$$\Delta Agg_t = c + \lambda_0 \Delta Agg_t^* + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \sum_{k=1}^{K} \lambda_k \Delta Agg_{t-k}^* + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t, \tag{VI.1}$$

where, $Agg_t$ is the aggregate; and $S_{it}$ represent centered seasonal dummies which take the value $1 - 1/12$ in the $i^{th}$ month and $-1/12$ otherwise. In this model, we select the regressors by *Autometrics* with Impulse Indicator Saturation and and apply the same two steps aforementioned procedure. We do not select over $Agg_t^*$, so that it is always present in the model. The maximum number of lags, K, is 4.

We compare the results of the indirect exercises between each other and with five direct alternatives. The direct procedures differ between each other on the potential regressors considered. Models may include only the aforementioned eight (seven) sub-aggregates (*cat 1*), a more detailed disaggregation of twenty four or twenty two broad categories, for the CPI and IPI, respectively (*cat 2*), or no disaggregated prices at all. Besides, when not including disaggregated prices, models may include the REER or not. Finally, we also consider a direct alternative based on Dynamic Factor Models. We extract the factors applying principal components to the first

difference of all the disaggregates, and include lags of this factor(s) as regressors. In all cases we apply the two steps procedure to select variables and impulses. The five direct exercises are summarized in table VI.2.

Table VI.1: Summary of the indirect forecasting exercises

| | Model | Description |
|---|---|---|
| 1 | **I** | Baseline disaggregate.<br>Individual univariate models for all the components<br>$\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12} \Delta x_{i,t-12} + \phi_{24} \Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ |
| 2 | **I -DFM-2** | Indirect procedure with two DF.<br>$\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k_1} + \phi_{12} \Delta x_{i,t-12} + \phi_{24} \Delta x_{i,t-24} +$<br>$\sum_{k=1}^{K} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ |
| | | **Pairwise** |
| 3 | **I - CC&CT** | Single-equation model for all the components<br>including CC and CT restrictions (when applicable)<br>**(i)** $\Delta x_{i,t} = c + \delta_i CC_{i,t-1} + \sum_{i=1}^{11} \gamma_i S_{i,t} + (\sum_{r=1}^{R} \alpha_{i,r} CR_{r,t-1}) + \epsilon_{i,t}$<br>**(ii)** $\Delta x_{i,t} = c + \sum_{r=1}^{R} \alpha_{i,r} CR_{r,t-1} + \sum_{k=1}^{K} \Delta\phi_k x_{i,t-k} + \phi_{12} \Delta x_{i,t-12} +$<br>$\phi_{24} \Delta x_{i,t-24} + \sum_{j=1}^{J} \theta_j \Delta SubAggCT_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$<br>**(iii)** $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12} \Delta x_{i,t-12} + \phi_{24} \Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ |
| 4 | **I - CC&CT-GP** | Single-equation model for all the components with CC or CT<br>restrictions. For the others, only its sub-aggregate is forecast |
| 5 | **I - onlyCT-N1** | Single-equation model for all the components including only CT<br>restrictions (when applicable). Normalization 1<br>**(i)** $\Delta x_{i,t} = c + \sum_{r=1}^{R} \alpha_{i,r} CR_{r,t-1} + \sum_{k=1}^{K} \Delta\phi_k x_{i,t-k} + \phi_{12} \Delta x_{i,t-12} + \phi_{24} \Delta x_{i,t-24}$<br>$+ \sum_{j=1}^{J} \theta_j \Delta SubAggCT_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$<br>**(ii)** $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ |
| 6 | **I - onlyCT-N2** | Single-equation model for all the components including only CT<br>restrictions (when applicable). Normalization 2 |
| 7 | **I - onlyCT-N1-GP** | Single-equation for all the components with only CT<br>restrictions. For the others, only its sub-aggregate is forecast |
| 8 | **I - onlyCC** | Single-equation model for all the components including only CC<br>restrictions (when applicable).<br>**(i)** $\Delta x_{i,t} = c + \sum_{s=1}^{S} \delta_{i,S} CC_{S_i,t-1} + \sum_{i=1}^{11} \gamma_i S_{i,t} + (\sum_{r=1}^{R} \alpha_{i,r} CR_{r,t-1}) + \epsilon_{i,t}$<br>**(ii)** $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ |
| 9 | **I - onlyCC-GP** | Single-equation model for all the components only CC<br>restrictions. For the others, only its sub-aggregate is forecast |

- All the equations represent the initial GUMs from where final models are selected using *Autometrics* with Impulse Indicator Saturation. The only variables on which we do not select ('unrestricted' regressor) are the constant and the common cycles — when it is the case. The selection is carried out in two steps. First we use a target size of 0.5% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.
- $K = J = 4$.
- In model 2 the factors are forecast in a VAR model, where lags are selected with *Autometrics* with IIS. The same two step procedure applies in this case.

Table VI.2: Summary of the direct forecasting exercises

| | Model | Description |
|---|---|---|
| 1 | **D** | Direct baseline.<br>Scalar model for the CPI<br>$\Delta Agg_t = c + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \phi_{12}\Delta Agg_{t-12} + \phi_{24}\Delta Agg_{t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$ |
| 2 | **D-DI-1** | Direct with disaggregated information 1.<br>$\Delta Agg_t = c + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \phi_{12}\Delta Agg_{t-12} + \phi_{24}\Delta Agg_{t-24} +$<br>$\sum_{i=1}^{8}\sum_{j=1}^{J} \theta_{i,j}\Delta SubAgg_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$ |
| 3 | **D-DI-2** | Direct with disaggregated information 2.<br>$\Delta Agg_t = c + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \phi_{12}\Delta Agg_{t-12} + \phi_{24}\Delta Agg_{t-24} +$<br>$\sum_{i=1}^{24}\sum_{j=1}^{J} \theta_{i,j}\Delta SubAgg_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$ |
| 4 | **D-REER** | Direct with REER.<br>$\Delta Agg_t = c + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \phi_{12}\Delta Agg_{t-12} + \phi_{24}\Delta Agg_{t-24} +$<br>$\sum_{q=1}^{Q} \delta_q \Delta REER_{t-q} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$ |
| 5 | **D-DFM-1** | Direct with 1 Factor.<br>$\Delta Agg_t = c + \sum_{k=1}^{K} \phi_k \Delta Agg_{t-k} + \phi_{12}\Delta Agg_{t-12} + \phi_{24}\Delta Agg_{t-24} +$<br>$\sum_{k=1}^{K} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$ |

- All the equations represent the initial GUMs from where models are selected using *Autometrics* with Impulse Indicator Saturation. The selection is carried out in two steps. First we use a target size of 0.5% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.
- $K = J = 4$ and $Q = 3$.
- In models 5 the factor is forecast in a scalar model, where lags are selected with *Autometrics* with IIS. The same two step procedure applies in this case.

## VI.4 Empirical results for the US CPI

Since the description of the data, the analysis of outliers, and the search of 'fully cointegrated' subsets for the US CPI was included in chapter III, here we focus just on the search of 'single-cycle' subsets and the forecasting exercise.

### VI.4.1 Common cycles tests

#### VI.4.1.1 Results of the pairwise tests: common cycles

Among the 164 components there are 13366 possible pairs. Without considering the aforementioned quality conditions (non zero coefficients and stability) we find 8846 pairs with a common cycle. After disregarding those that do not satisfy the the two conditions we keep 2783 pairs that are considered to satisfy stable and statistically sensible restrictions.

Table VI.3 summarizes the results of the pairwise procedure with common cycles. It contains the amount and weight of all the series included in some 'single-cycle' subset.

Table VI.3: Summary of the pairwise procedure results with common cycles: total proportion of series and weight of all the blocks (with all CIrels, all controls and *sequential* $REER^{(a)}$)

Total proportion of series and weight of all the *Almost fully connected sets*

| Num of sets | Num of Sers | | Weight(%)* | | Weight top three(%)** |
|---|---|---|---|---|---|
| | CC | $CC\&CT^{(b)}$ | CC | $CC\&CT$ | |
| 15 | 92 | 28 | 30.0 | 5.9 | 7.0 |

*(a)* In this strategy blocks are firstly formed without *REER* and then the procedure with *REER* is executed only for the series not included in any block.
*(b)* This column contains the number of series that belong to some almost fully connected set *and* some almost fully cointegrated set.
(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the CPI.
(**) Weight of the tree heavier series.

Table VI.4 gives some further details. Conclusions are similar to those from table III.8. Blocks of series sharing one common cycle (or two for groups with *REER*) cannot be assigned to a single broad category. However, in almost all the cases, more than 85% of the *Set*'s weight is explained by two broad categories. The exceptions are blocks 3 to 6 for which the two most important categories explain more than 65% of the block's weights.

Table VI.4: Detailed results of the Pairwise procedure: number of series and proportion of weight by broad categories and blocks

| | MAN dur | | PF | | SERV | | NPF | | MAN No dur | | ENE | | Tot W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q | W | Q | W | Q | W | Q | W | Q | W | Q | W | |
| TOTAL | 51 | 17.4 | 38 | 12.3 | 34 | 51.8 | 25 | 4.6 | 10 | 3.7 | 6 | 10.1 | |
| Block 1 | 5 | 57.7 | 1 | 1.6 | 1 | 4.1 | 2 | 6.4 | 1 | 30.2 | 0 | 0.0 | 4.7 |
| Block 2 | 6 | 66.2 | 2 | 6.8 | 2 | 27.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1.8 |
| Block 3 | 2 | 6.1 | 3 | 19.2 | 2 | 57.1 | 0 | 0.0 | 2 | 17.6 | 0 | 0.0 | 2.2 |
| Block 4 | 3 | 44.9 | 2 | 15.3 | 2 | 13.6 | 2 | 26.2 | 0 | 0.0 | 0 | 0.0 | 1.9 |
| Block 5 | 4 | 33.7 | 1 | 4.3 | 1 | 37.8 | 2 | 24.2 | 0 | 0.0 | 0 | 0.0 | 1.8 |
| Block 6 | 2 | 17.6 | 1 | 5.4 | 2 | 37.8 | 2 | 12.2 | 2 | 27.1 | 0 | 0.0 | 2.9 |
| Block 7 | 1 | 10.4 | 2 | 10.2 | 2 | 79.4 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1.3 |
| Block 8 | 3 | 81.5 | 0 | 0.0 | 0 | 0.0 | 1 | 18.5 | 0 | 0.0 | 0 | 0.0 | 1.2 |
| Block 9 | 1 | 39.4 | 0 | 0.0 | 0 | 0.0 | 3 | 60.6 | 0 | 0.0 | 0 | 0.0 | 0.5 |
| Block 10 | 1 | 2.6 | 0 | 0.0 | 2 | 86.7 | 1 | 10.7 | 0 | 0.0 | 0 | 0.0 | 2.6 |
| Block 11 | 2 | 38.7 | 2 | 61.3 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1.0 |
| Block 12 | 0 | 0.0 | 3 | 47.8 | 1 | 52.2 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1.7 |
| Block 13 | 2 | 26.3 | 0 | 0.0 | 1 | 71.4 | 1 | 2.3 | 0 | 0.0 | 0 | 0.0 | 2.5 |
| Block 14 | 2 | 6.4 | 0 | 0.0 | 0 | 0.0 | 1 | 2.4 | 0 | 0.0 | 1 | 91.2 | 3.5 |
| Block 15 | 0 | 0.0 | 1 | 18.5 | 2 | 66.5 | 1 | 14.9 | 0 | 0.0 | 0 | 0.0 | 0.5 |
| TOTAL | 34 | | 18 | | 18 | | 16 | | 5 | | 1 | | 30.0 |

Columns $Q$ indicate the amount of series in each category and Set.
Columns $W$ indicate the the total weight of each category in the CPI and the proportion of the weight of each category in each Set.
Last column contains the total weight of the blocks.

### VI.4.2 Forecasts for all the US CPI components with the I-PW approach

Figure VI.1, figure VI.2, table VI.5, and VI.6 summarize the detailed analysis of the components. Figure VI.1 includes all the components' forecasts in a single plot, and figure VI.2 includes the box plots of the 164 components for each of the 24 months of years 2013 and 2014. From these two figures, two observations can be drawn: (1) The number of components outside the whiskers is larger in the forecast period than in 2013. (2) In 2013 the headline inflation seems to be around the medians of the components forecasts, but in 2014 there is a greater proportion of components forecasts below the headline forecasts, meaning that components with greater weights in the CPI are forecast with greater rates of growth.

Table VI.5 gives disaggregated forecasts for the average annual growth rates of the US CPI and its components for 2014 made with information up to December 2013. The table uses green shadows to indicate that the point forecasts of the components are below the lower bound of the confidence interval for CPI and red for the components' forecasts above the corresponding upper

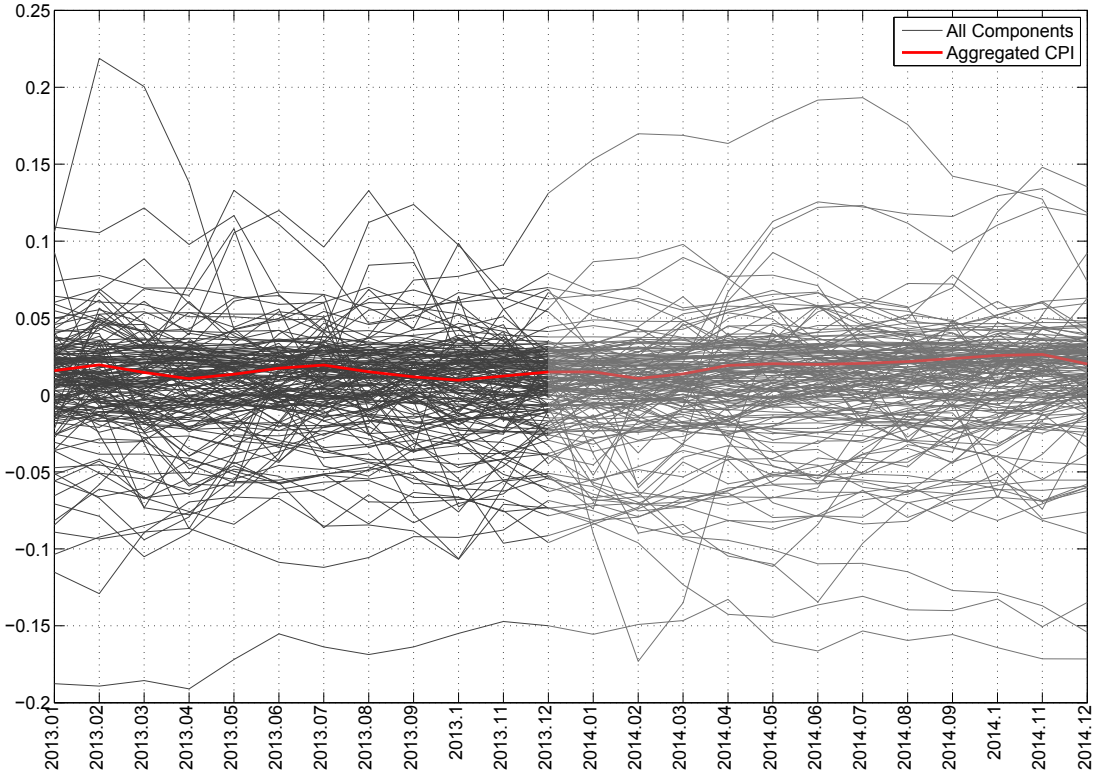
bound. Italics and bold letters are used to indicate that the weight of a particular component is relatively high (see the *Notes to table VI.5* below for a detailed description).

Table VI.5 shows that the components' forecasts below the mentioned lower bound belong mainly to the category of non-energy industrial goods. Components' forecasts above the upper bound correspond mainly to the other categories, having special impact in the headline inflation the prices in the energy group. The services which were forecast to be the most inflationary ones for 2014 are: *Technical and business school tuition and fees*, *Funeral expenses*, *Rent of primary residence*, *Water and sewerage maintenance*, *Dental services*, *Hospital services*, *Nursing homes and adult day services*, *Pet services including veterinary*, *Delivery services*, *Repair of household items*, *College tuition and fees*, *Elementary and high school tuition and fees*, *Postage*, and *Housing at school excluding board*. On the contrary, the services with negative inflation forecasts are *Airline fare, Wireless telephone services*, and *Health insurance*.

An important number of components in the group of non-energy industrial goods show expected negative inflation, being especially negative the expectations in *Men's suits, sport coats, and outerwear*, *Personal computers and peripheral equipment*, *Computer software and accessories*, *Telephone hardware, calculators, and other consumer information items*, *Other linens*, *Clocks, lamps, and decorator items*, *Televisions*, *Other video equipment*, *Audio equipment*, *Photographic equipment and supplies*, and *Toys*.

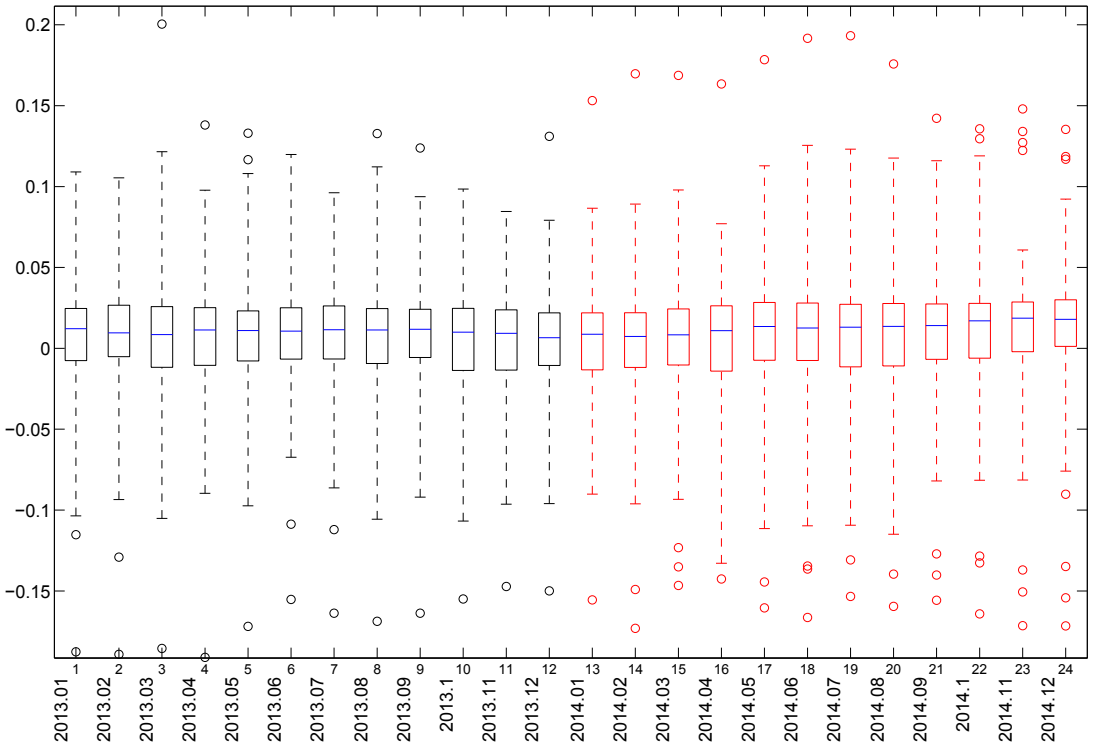
Table VI.6 classifies all the components according to our ability to forecast them. Red colors are used for components' which are relatively hard to forecast (large Root Mean Square Forecast Errors) and green for relatively easy to forecast ones. Additionally, highly inflationary components (those with red color in column *2014* of table VI.5) are marked with a dotted shadow (see the *Notes to table VI.6* below for a detailed description).

Three main conclusions can be extracted from table VI.6: (i) almost all energy and non-processed food components are relatively hard to forecast, (ii) almost all services are relatively easy to forecast, (iii) in non-energy industrial goods and processed food there is not a clear pattern.




Dark lines are observed values and light ones are forecasts.

Figure VI.1: Components' forecasts at 2013.12 ($\Delta_{12} logP_t$)




- The edges of the box are the $25^{th}$ ($q_1$) and $75^{th}$ ($q_3$) percentiles.
- The whiskers extend from $q_1 - 3(q_3 - q_1)$ to $q_3 + 3(q_3 - q_1)$.
- Black color indicates observed values, red, forecasts.

Figure VI.2: Box plots of the observed and forecast components at 2013.12 ($\Delta_{12} logP_t$)

**Notes to table VI.5**

- Columns *09.13* include the mean of annual growth of each log price ($\Delta_{12} logP_t$) in December of years 2009 to 2013.
- Columns *2014* include the forecast of $\Delta_{12} logP_{2014.12}$ with information up to December 2013.
- Names Reference:
  - Man Dur: Non energy industrial durable goods.
  - SERV: Non energy services.
  - EN: Energetic goods and services.
  - NPF: Non-processed food.
  - M_N: Industrial non-durable goods.
  - PF: Processed food.
- Color references for columns *09-13*
  - Green: the forecast for $\Delta_{12} logP_{2014.12}$ is smaller than the lower bound of the 95% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2009 to 2013.
  - Red: the forecast for $\Delta_{12} logP_{2014.12}$ is larger than the upper bound of the 95% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2009 to 2013.
  - White: the forecast for $\Delta_{12} logP_{2014.12}$ is inside the 95% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2009 to 2013.
  - Standard deviation for the mean of ($\Delta_{12} logP_t$) at December of years 2009 to 2013 is computed as $\frac{\sigma}{\sqrt{6}}$ where $\sigma$ is the sample standard deviation $\Delta_{12} logP_t$. Autocovariances at lags 12 and larger are ignored, and the confidence interval is constructed assuming normality.
- Color references for columns *2014*
  - Green: the forecast at December 2013 for $\Delta_{12} logP_{2014.12}$ is smaller than the lower bound of the 95% confidence interval for the forecast of $\Delta_{12} logCPI_{2014.12}$.
  - Red: the forecast at December 2013 for $\Delta_{12} logP_{2014.12}$ is larger than the upper bound of the 95% confidence interval for the forecast of $\Delta_{12} logCPI_{2014.12}$.

- White: the forecast at December 2013 for $\Delta_{12} logP_{Dec14}$ is inside the 95% confidence interval for the forecast of $\Delta_{12} logCPI_{2014.12}$.
- Blue italics: indicate component's weights larger than the average weight $(1/N)$
- Blue bold: indicate component's weights larger than the 3%.
- Standard deviation for the 12 step ahead forecast error of $\Delta_{12} logCPI_t$ is computed as the historical out of sample Root Mean Squared Forecast Error (see table VI.7).
- Grey shadows in components' names are just to distinguish between categories.

Table VI.5: Components' forecasts ($\Delta_{12} log P_t$)

| MAN dur | 09-13 | 2014 | SERV | 09-13 | 2014 | EN/NPF/M_N | 09-13 | 2014 | PF | 09-13 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Men's suit | -1.5 | -9.0 | Technical | 3.9 | 4.5 | Fuel oil | 3.6 | 11.9 | Flour and | 5.4 | 2.8 |
| Men's furn | 3.8 | 1.4 | Telephone | 0.6 | *0.0* | Propane, k | -0.6 | 7.4 | Fresh bisc | 3.9 | 3.2 |
| Men's shir | 0.1 | -1.7 | Internet s | 0.7 | *0.3* | Electricit | 2.0 | *1.4* | Cakes, cup | 3.3 | 0.9 |
| Men's pant | 1.0 | -0.4 | Haircuts a | 1.5 | *2.0* | Utility (p | -4.8 | *-2.1* | Other bake | 3.4 | 2.8 |
| Boys' appa | 1.7 | -0.5 | Legal serv | 3.0 | 3.3 | Gasoline ( | 2.1 | **13.5** | Pork | 2.4 | 3.4 |
| Women's ou | 1.9 | -3.4 | Funeral ex | 3.0 | 3.5 | Other moto | 2.7 | 11.7 | Bacon, bre | 2.8 | 3.0 |
| Women's dr | -0.4 | 1.8 | Laundry an | 2.5 | 2.7 | Breakfast | 1.8 | 1.1 | Ham | 2.1 | 6.0 |
| Women's su | -0.8 | *-0.9* | Apparel se | 4.0 | 2.4 | Rice, past | 5.2 | 3.2 | Pork chops | 2.0 | 1.4 |
| Women's un | 2.7 | 0.5 | Financial | 1.3 | 3.0 | Bread | 3.5 | 2.4 | Other poul | 5.1 | 3.8 |
| Girls' app | 0.5 | -0.2 | Rent of pr | 2.0 | **3.3** | Uncooked g | 5.4 | 4.6 | Processed | 4.0 | 2.9 |
| Men's foot | 2.3 | 0.4 | Other lodg | -1.3 | *1.4* | Uncooked b | 4.5 | 4.0 | Ice cream | 2.6 | 2.1 |
| Boys' and | 2.3 | 1.8 | Owners' eq | 1.4 | **2.8** | Uncooked b | 3.2 | 2.7 | Other dair | 1.2 | 1.7 |
| Women's fo | 1.3 | 0.2 | Tenants' a | 2.7 | 2.1 | Uncooked o | 5.6 | 3.7 | Canned fru | 4.1 | 2.2 |
| Infants' a | 1.0 | -0.8 | Water and | 6.3 | *4.4* | Other pork | 2.3 | 3.2 | Frozen fru | 2.5 | 3.4 |
| Watches | 1.1 | -0.2 | Garbage an | 3.0 | 3.2 | Other meat | 2.5 | 2.9 | Other proc | 4.2 | 2.7 |
| Jewelry | 4.1 | 1.6 | Domestic s | 1.2 | 2.2 | Chicken | 2.6 | 3.3 | Carbonated | 2.8 | 2.9 |
| Personal c | -10.6 | -15.4 | Moving, st | 0.1 | 1.0 | Fresh fish | 3.2 | 3.4 | Frozen non | 3.1 | 2.2 |
| Computer s | -4.7 | -6.0 | Physicians | 2.7 | *2.2* | Eggs | -0.6 | 4.3 | Nonfrozen | 0.7 | 0.4 |
| Telephone | -4.0 | -7.6 | Dental ser | 2.9 | *3.7* | Milk | 0.2 | 0.7 | Other beve | 0.8 | 1.2 |
| Floor cove | -1.6 | 0.5 | Services b | 1.9 | 2.3 | Cheese and | 2.0 | 1.2 | Sugar and | 3.8 | 0.7 |
| Window cov | -2.7 | -2.3 | Hospital s | 6.2 | *6.3* | Apples | 3.3 | -1.5 | Candy and | 3.4 | 2.6 |
| Other line | -5.0 | -6.2 | Nursing ho | 3.2 | 3.7 | Bananas | 2.3 | 1.9 | Other swee | 3.9 | 2.4 |
| Bedroom fu | -0.9 | 0.3 | Pet servic | 4.1 | 4.7 | Citrus fru | 1.6 | 3.9 | Butter and | 5.6 | 3.9 |
| Living roo | -0.2 | -1.2 | Club dues | 0.3 | *1.3* | Other fres | -1.4 | 6.3 | Salad dres | 4.0 | 0.9 |
| Other furn | -1.3 | -2.7 | Admissions | 1.8 | *2.7* | Potatoes | 1.2 | 3.4 | Other fats | 6.2 | 2.7 |
| Major appl | 0.3 | -0.6 | Fees for l | 2.2 | 2.9 | Lettuce | 0.0 | 9.2 | Soups | 1.5 | 1.7 |
| Other appl | -1.4 | -2.4 | Motor vehi | 2.5 | 2.6 | Tomatoes | -2.7 | 2.7 | Frozen and | 1.5 | 0.5 |
| Clocks, la | -5.0 | -5.5 | Motor vehi | 2.8 | 2.7 | Other fres | 1.1 | 1.5 | Snacks | 4.9 | 3.4 |
| Dishes and | -4.3 | -3.9 | Motor vehi | 2.6 | *2.8* | Coffee | 3.8 | 0.1 | Spices, se | 3.0 | 1.8 |
| Nonelectri | 0.7 | 0.9 | State moto | 3.6 | 1.2 | Tobacco pr | 7.1 | 4.8 | Baby food | 2.6 | 2.3 |
| Tools, har | 0.1 | 0.4 | Parking an | 4.3 | 3.2 | Hair, dent | -0.3 | -0.1 | Other misc | 2.3 | 1.1 |
| Outdoor eq | -0.8 | -0.9 | Airline fa | 3.6 | *-2.0* | Cosmetics, | 1.0 | 0.6 | Full servi | 2.4 | *2.2* |
| Eyeglasses | 0.7 | 1.8 | Other inte | -0.2 | 0.1 | Miscellane | -0.3 | -1.2 | Limited se | 2.8 | *2.7* |
| Television | -22.6 | -17.2 | Intracity | 4.2 | 2.6 | Household | 1.8 | 0.4 | Food from | 3.3 | 2.1 |
| Cable and | 2.3 | *3.4* | Child care | 3.1 | *2.6* | Household | 3.9 | 3.1 | Other food | 2.8 | 2.8 |
| Other vide | -13.0 | -13.5 | Delivery s | 6.9 | 4.1 | Miscellane | 1.1 | 1.3 | Beer, ale, | 2.4 | 2.2 |
| Video disc | -0.1 | -1.1 | Wireless t | -1.5 | *-0.7* | Prescripti | 3.1 | *3.9* | Distilled | 0.9 | 1.2 |
| Audio equi | -5.4 | -5.8 | Food at em | 3.5 | 3.0 | Photograph | 2.0 | 0.6 | Wine at ho | 0.4 | 0.9 |
| Audio disc | -3.4 | -1.2 | Gardening | 0.9 | 2.3 | Newspapers | 3.7 | 2.3 | Alcoholic | 3.0 | 3.2 |
| Pets and p | 3.1 | *1.1* | Repair of | 4.1 | 4.4 | Recreation | -0.8 | -0.3 | **PF** | | **2.3** |
| Sports veh | 1.6 | 1.5 | Land-line | 2.0 | *1.0* | Nonprescri | -0.1 | -0.3 | | | |
| Sports equ | -0.9 | -1.4 | Care of in | 1.5 | 1.0 | **ENE** | | **8.1** | | | |
| Photograph | -5.5 | -6.0 | Health ins | 0.9 | *-1.3* | **NPF** | | **2.8** | | | |
| Toys | -5.8 | -5.9 | College tu | 5.0 | *4.5* | **ManNd** | | **1.7** | | | |
| Sewing mac | 3.0 | 0.7 | Elementary | 4.1 | 3.4 | | | | | | |
| Music inst | 0.2 | -0.1 | Postage | 3.4 | 3.3 | | | | | | |
| New vehicl | 1.2 | **0.1** | Housing at | 4.3 | 3.7 | | | | | | |
| Used cars | 1.2 | *-4.5* | **SERV** | | **2.6** | | | | | | |
| Car and tr | 1.9 | -0.7 | | | | | | | | | |
| Tires | 3.5 | 2.2 | | | | | | | | | |
| Vehicle ac | 3.9 | 3.0 | | | | | | | | | |
| Indoor pla | 0.3 | -2.3 | | | | | | | | | |
| Medical eq | 0.5 | 0.9 | | | | | | | | | |
| Leased car | -1.1 | -1.6 | | | | | | | | | |
| **MAN_Dur** | | **-1.1** | | | | | | | | | |

| | CPI | Fore(h12) |
|---|---|---|
| **2013** | **1.5** | **2.0** |
| **2014** | **0.8** | **2.1** |

2014

Fore < CPI - 0.98
Fore = CPI +- 0.98
Fore > CPI + 0.98
*italic* Weight > 100/181
*bold* Weight larger than 3

**Confidence 95%**

2008-2013

Fore 14 < Mean(08-13)
Fore 14 = Mean(08-13)
Fore 14 > Mean(08-13)

**Notes to table VI.6**

- The table classifies all the components according to our ability to forecast them.
- Columns *h1*, *h6* and *h12* refer to forecasts horizons 1, 6 and 12 respectively
- Using all the components' RMSFE for each forecast horizon we compute the quintiles and classify the components according to the quintile to which they belong.
- Names Reference:
  - Man Dur: Non energy industrial durable goods.
  - SERV: Non energy services.
  - EN: Energetic goods and services.
  - NPF: Non-processed food.
  - M_N: Industrial non-durable goods.
  - PF: Processed food.
- Color reference:
  - Dark Red: $Q4 \leq RMSFE_i < Q5$. (relatively 'hard' to forecast).
  - Light Red: $Q3 \leq RMSFE_i < Q4$.
  - White: $Q2 \leq RMSFE_i < Q3$.
  - Dark Green: $Q1 \leq RMSFE_i < Q2$.
  - Light Green: $RMSFE_i < Q1$ (relatively 'easy' to forecast).
  - Grey shadows in components' names are just to distinguish between categories.
- In addition to the color classification we fill with dots those components for which the forecast at December 2013 for $\Delta_{12} logP_{2014.12}$ is larger than the upper bound of the 95% confidence interval for the forecast of $\Delta_{12} logCPI_{2014.12}$ (red components in table VI.5).

Table VI.6: Components RMSFE

### VI.4.3 Comparative forecasting results

For a complete description of the forecasting exercises see §VI.3. Table VI.7 includes a summary of the forecasting results. The first row includes the root mean squared forecast error in the direct baseline for $\Delta_{12} log(CPI)$ for horizons $H = 1$ to $H = 12$. All the other entries in the table are ratios with respect to the baseline. The evaluation period is $2011.1 - 2014.12$. This evaluation period may seem rather short, specially when analyzing medium term RMSFEs. For example, for $H = 12$ we have thirty six forecasts, but only three of them are independent. Therefore, results for horizons larger than four or five steps ahead must be interpreted with extreme caution.

The selection of the sample size was related to the availability of disaggregated data. We could have chosen a smaller number of disaggregates, extend our sample from the beginning, and enlarge the evaluation period of the forecasting exercise. As discussed in the introduction, our procedure requires working with high levels of disaggregation since intermediate sub-aggregates may add up components that do not share common common trends or cycles, precluding the formation of the subsets we are looking for. In the trade-off between more disaggregation vs. larger evaluation period, we gave more importance to the former. Although the maximum publicly available disaggregation level — the one we are considering — still contains aggregated series, this is the best we can do.

In our second application we solve this problem by considering larger samples (see §VI.5).

#### Direct approaches

The use of the Real Effective Exchange Rate (option *D-REER*) does not seem to make any relevant difference.

*Disaggregated information (D-DI-1 and D-DI-2)*

When using disaggregated information in a scalar model for the aggregate, as proposed by Hendry and Hubrich (2011), low disaggregation levels are preferred to higher ones[1]. For the lowest disaggregation level (8 sub-aggregates), while results are somewhat worst than in the baseline for short horizons, in longer ones (horizons 10 to 12) the use of disaggregated information improves over the baseline.

[1] In table 6 of Hendry and Hubrich (2011) there is another example of how the forecast accuracy deteriorates as the level of disaggregation increases.

*Scalar models enlarged with Dynamic Factors (D-DFM-1)*

The best results are obtained with only one factor. The inclusion of the factor produces some gains over the baseline in all horizons except in 7 to 9, where results are indistinguishable.

**Indirect approaches**

As a brief summary; we found that the baseline option is hard to beat in short horizons (1-5). Although when adding a dynamic factor, *D-DFM-1*, there are some improvements in the forecasting accuracy, they are only minor with ratios between 0.99 and 0.96.

Simple indirect methods perform poorly in short horizons. However, this bad performance can be corrected by the consideration of short run restrictions between the components. In longer horizons disaggregation becomes very relevant with RMSFE being 60% of the baseline in the basic indirect approach. This good performance can still be somehow improved by the inclusion of long run restrictions between the components.

*Univariate models for all the components (I)*

This is the simpler disaggregated approach. In short horizons (1-6) it clearly deteriorates with respect to the baseline. From horizons 7 to 12 this approach considerably outperforms the the baseline. In $H = 12$ the RMSFE is 60% of the baseline.

*Univariate models for all components enlarged with Dynamic Factors (I-DFM-2)*

The best results are obtained with two factors. The advantage of the indirect forecast completely disappears if univariate models are enlarged with Dynamic Factors, and the bad performance for short horizons is even worst. In this application the use of DFM to forecast the components is not useful at all, univariate models do much better.

*Pairwise procedures*

In these procedures the basic components that do not belong to a subset sharing one common trend, one common cycle or both, are forecast using univariate individual models. An alternative consists in forecasting just the aggregate of those basic components ($I - PW - GP$). In short horizons the latter alternative provides slightly better forecast for the aggregated CPI than the corresponding case in which basic components outside subsets with common features are forecast individually. For longer forecast the conclusion is the opposite.

The use of different normalizations has little effect in the RMSFE. The largest difference between the two possibilities is 4 points (at horizons 7 and 8).

When considering only common unit roots restrictions results are better than in the basic indirect procedure (I). In particular the approach *I-onlyCT-N1-GP* clearly dominates *I* in the first four horizons[2]. However when considering only common unit roots, indirect procedures are still worse than direct ones in the initial horizons. In long horizons, when the basic indirect clearly dominates the baseline, the inclusion of common unit roots delivers some further improvements.

If instead of common unit roots we consider only common cycles, the improvements in short horizons with respect to I are even larger and results are close to the baseline (except for H=1). In larger horizons results are similar to the basic indirect.

Previous comments are also valid for *I-CC&CT* as results are almost indistinguishable form *I-CC*.

Interestingly enough we found that when not using IIS in the forecasting equations, the indirect procedures strongly deteriorate. This highlight the importance of the outliers and breaks treatment when dealing with disaggregates. In contrast, the direct procedure is not highly affected by the use of IIS.

As a conclusion, this exercise shows that the modeling of the basic components by single-equation methods taking into account the common features restrictions between them, identified by pairwise methods, can be considered as adequate in the sense that the forecasts for the aggregate are quite good. This is an interesting indirect test of the common-feature disaggregated approach.

As we mentioned in chapter III, the components tend to be more contaminated with outliers than the aggregate, this necessarily affect the forecasts, and the pernicious effects could be more relevant for the pairwise procedure. Therefore, the application of robustifying procedures as the proposed by Hendry (2006) and studied by Castle et al. (2015) could be interesting.

[2]Recall that the total weight of the series in some set with common unit roots is only 11%.

Table VI.7: Relative RMSE $\Delta_{12}log(CPI)$. Evaluation period: 2011.1 - 2014.12 (first row: RMSE for the baseline. All the others are ratios with respect to the first)

| | H=1 | H=2 | H=3 | H=4 | H=5 | H=6 | H=7 | H=8 | H=9 | H=10 | H=11 | H=12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D** | **0.22** | **0.39** | **0.50** | **0.52** | **0.56** | **0.59** | **0.67** | **0.74** | **0.80** | **0.83** | **0.88** | **0.96** |
| D-DI-1 | 1.03 | 1.02 | 1.05 | 1.09 | 1.10 | 1.04 | 1.02 | 1.04 | 1.01 | 0.97 | 0.92 | 0.91 |
| D-DI-2 | 1.24 | 1.20 | 1.19 | 1.19 | 1.16 | 1.21 | 1.23 | 1.18 | 1.10 | 1.11 | 1.15 | 1.23 |
| D-REER | 1.00 | 0.99 | 0.99 | 0.98 | 0.97 | 0.98 | 1.00 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 |
| D-DFM-1 | 0.98 | 0.97 | 0.96 | 0.97 | 0.98 | 0.99 | 1.01 | 1.00 | 1.00 | 0.98 | 0.96 | 0.97 |
| I | 1.15 | 1.12 | 1.12 | 1.14 | 1.12 | 1.09 | 1.03 | 0.94 | 0.83 | 0.71 | 0.65 | 0.59 |
| I-DFM-2 | 1.47 | 1.34 | 1.27 | 1.29 | 1.26 | 1.25 | 1.20 | 1.21 | 1.19 | 1.12 | 1.07 | 1.01 |
| **Pairwise** | | | | | | | | | | | | |
| I-CC&CT | 1.07 | 0.99 | 0.99 | 1.05 | 1.01 | 0.99 | 0.95 | 0.90 | 0.82 | 0.72 | 0.67 | 0.62 |
| I-CC&CT-GP | 1.04 | 0.96 | 0.96 | 1.02 | 1.03 | 0.99 | 0.93 | 0.87 | 0.82 | 0.74 | 0.71 | 0.66 |
| I-onlyCT-N1 | 1.09 | 1.06 | 1.10 | 1.14 | 1.10 | 1.08 | 0.99 | 0.90 | 0.81 | 0.70 | 0.63 | 0.59 |
| I-onlyCT-N2 | 1.11 | 1.07 | 1.09 | 1.13 | 1.11 | 1.11 | 1.03 | 0.94 | 0.84 | 0.72 | 0.64 | 0.59 |
| I-onlyCT-N1-GP | 1.05 | 1.02 | 1.04 | 1.08 | 1.10 | 1.09 | 1.01 | 0.95 | 0.93 | 0.91 | 0.89 | 0.87 |
| I-onlyCC | 1.06 | 0.98 | 0.99 | 1.04 | 1.01 | 0.99 | 0.94 | 0.89 | 0.81 | 0.72 | 0.67 | 0.63 |
| I-onlyCC-GP | 1.07 | 1.00 | 1.02 | 1.09 | 1.11 | 1.08 | 1.01 | 0.93 | 0.87 | 0.81 | 0.78 | 0.74 |

- See tables VI.1 and VI.2 for a description of each exercise.
- Dark red entrances highlight the loser procedure.
- Light red indicates procedures that are, at most, 5 points smaller than the worst one.
- Dark green indicates the best procedure.
- Light green indicates procedures that are, at most, 5 points larger than the best one.

## VI.5 Empirical results for the US IPI

### VI.5.1 Data

The break down of the Industrial Production Index (IPI) that we use in this analysis corresponds to the maximum disaggregation level available to the public in the Board of Governors of the Federal Reserve System; 134 components for the period Jan-1980 to Dec-2014 (420 observations). We use the seasonal adjusted series provided by the board. This decision is based on two facts. First, the general assumption of linear growth and seasonal dummies, that was valid for prices, does not seem so for the IPI, as some disaggregates show seasonal unit roots. Second, the weights of the individual components are publicly available only for the seasonally adjusted series, so that we do not have other alternative than use those series. Although the use of seasonally adjusted series is known to generate some problems for econometric analysis, its use is quite common in empirical applications, see e.g.,Hendry and Hubrich (2011); Stock and Watson (2007); and Trenkler et al. (2007).

The use of seasonally adjusted series raises the issue highlighted by Cubadda (1999). He

argues that the spectral density matrix of $\Delta X_t$ cannot be null at any frequency under the presence of *strong serial correlation common features* (SCCF), and given that seasonal filters eliminate the spectrum at the seasonal frequencies, seasonally adjusted series should not be used in common cycles analysis. This problem do not affect affect our application as we are performing WF, not SCCF tests.

Another difference with respect to the CPI is the presence of calendar effects. The Board of Governors of the Federal Reserve System provides a list of twenty one possible calendar effects, but not all of them are necessarily relevant for all the components. Our modeling strategy is to include all the calendar effect in all the components' equations and select the relevant ones with *Autometrics*.

Figure VI.3 describes all the 134 components and the aggregated IPI. The first panel shows a great heterogeneity among the components, what could be an indicator of the importance of disaggregated analysis regardless of whether the interest is in the components or on the aggregate.

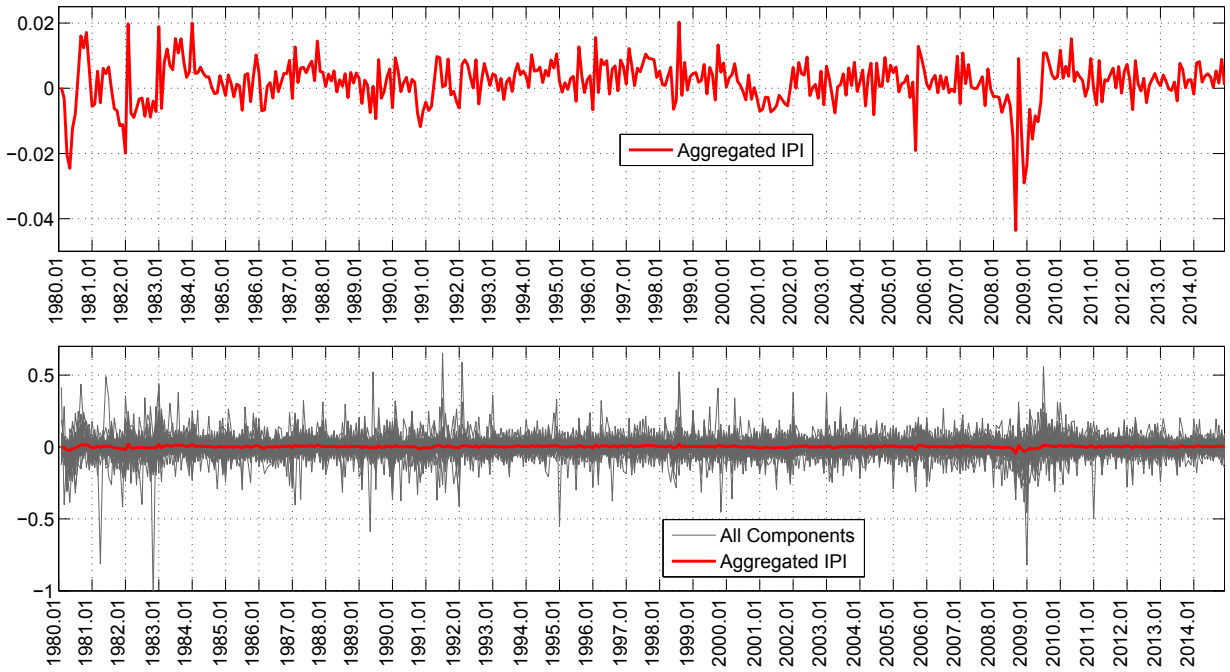



Figure VI.3: $\Delta logIPI$ and its components, 1980.1 - 2014.12

### VI.5.2 Outliers' analysis

As described in §III.2.2 we carry out the outliers search for the 134 components in individual models for the first differences using *Autometrics* with Impulse Indicator Saturation (IIS). We

select the impulses in two steps. First we use a target gauge of 0.5% to select lags, calendar effect and impulses, and store the retained impulses. In a second step we consider the same original model augmented with the retained impulses and a target size of 5% with no IIS.

To make tables legible, we group basic components into six broad categories: manufactured durable goods; food, beverages and tobacco; Chemicals; Other manufactured non-durable goods; Minerals; and Utilities.

Table VI.8 summarizes the results of the outliers' search. Four main observations emerge from it: i) the average number of outliers per series is 17.3 (4.1% of the observations); ii) durable goods and minerals are the most contaminated with a mean proportion of 4.8% and 4.7% of outlying observations per component, respectively; iii) 13% + 18% = 31% of the outliers are large (larger than $4\sigma$ in absolute value); and iv) minerals have the largest proportion of large outliers (37% of the total number of outliers).

The comparison between the analysis of outliers for the $IPI$ and the $CPI$ (see table III.4) leads to the following conclusions: (a) The mean proportion of outliers is larger in the IPI than in the CPI — 4.1% in the IPI *vs.* 2.6% of the observations in the CPI. (b) The proportion of outliers in the most contaminated groups are similar — 4.8% of the observations of the industrial production of durable goods, *vs.* 4.5% in energy prices. (c) Though somewhat higher in the CPI, the proportion of large outliers is similar — 31% of the total number of outliers in the IPI, *vs.* 37% in the CPI.

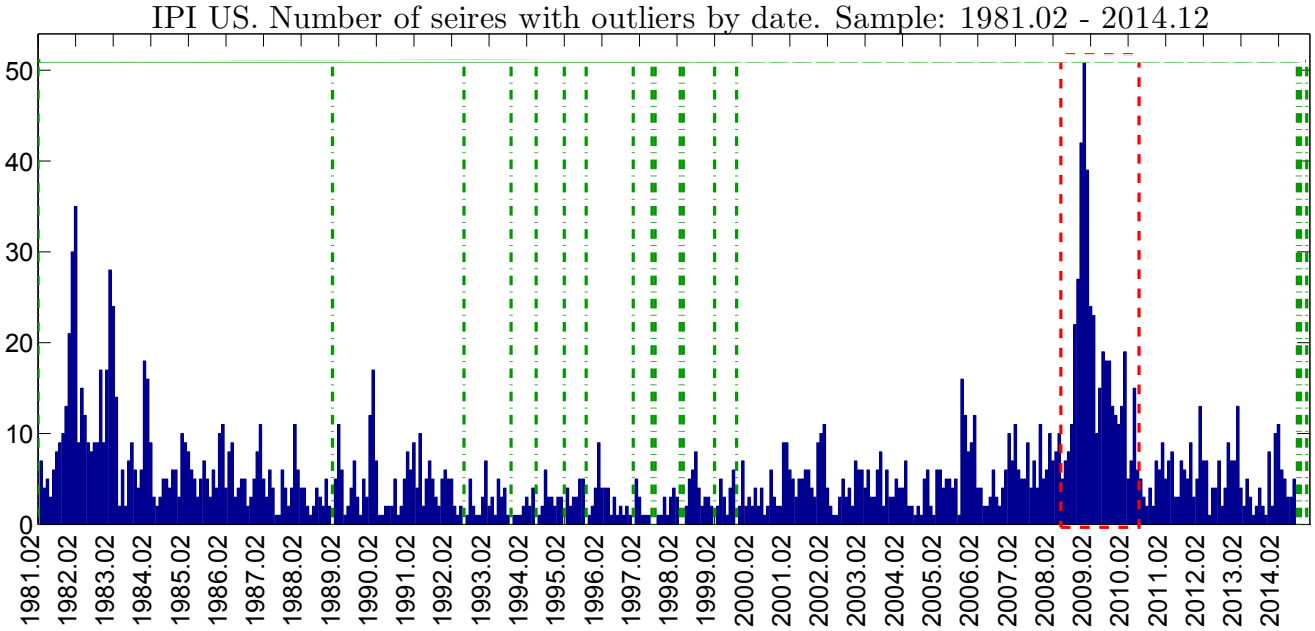
Table VI.8: Mean number of outliers by size and category

| | L+ | S+ | S- | L- | **Mean** | **Mean (% of T)** |
|---|---|---|---|---|---|---|
| MAN_Dur (63) | 2.9 | 6.0 | 7.4 | 3.9 | **20.2** | **4.8%** |
| Food,Bev,Tob (17) | 0.9 | 3.8 | 4.9 | 1.4 | **11.0** | **2.6%** |
| Chemicals (11) | 2.3 | 5.5 | 5.4 | 3.5 | **16.5** | **3.9%** |
| Oth NonDur Man (26) | 1.2 | 4.3 | 5.9 | 2.3 | **13.8** | **3.3%** |
| Minerals (7) | 3.6 | 5.3 | 7.3 | 3.7 | **19.9** | **4.7%** |
| Utilities (10) | 2.8 | 6.3 | 5.9 | 3.4 | **18.4** | **4.4%** |
| TOTAL (134) | 13.7 | 31.2 | 36.8 | 18.2 | **17.3** | **4.1%** |
| **PROP** | **13%** | **31%** | **38%** | **18%** | **100%** | |

Numbers in parenthesis after the category name indicate the number of series in the category.
L+: Large (larger than $4\sigma$) and positive outliers.
S+: Small (smaller than or equal to $4\sigma$) and positive outliers.
L-: Large and negative outliers.
S-: Small and negative outliers.

We also study the distribution of outliers by date. Figure VI.4 shows the number of series

with outliers at each of the 420 months of the sample. As it shows, the distribution by dates if far form uniform, with some months having 50 (37%) of the series with outliers and some others with none. Interestingly, as it was also the case of the CPI, there seems to be a concentration around years 2008-2009, the sub-prime crisis period (red box of the figure). Additionally, during the 'great moderation' period starting in the mid eighties and ending in the late nineties, there seems to be a reduction in the number of components with outliers.




Green doted lines indicate dates in which there are no contaminated series.

Figure VI.4: Distribution of series with outliers by date

As we did for the CPI, to study in more detail the distribution of the number of components with outliers, we estimate a model for the proportion of series with outliers including as potential regressors, seasonal dummies, the twenty one calendar effects, and choosing the *Autometrics* option IIS+SIS for outliers and breaks detection with a target gauge equal to $1/T$ . Results are summarized in figure VI.5, from where four important conclusions can be drawn (for the full model see appendix D.1 ): (i) as was the case for the CPI, in January there are, on average, more series with outliers; (ii) there is a relevant calendar effect (a variable that indicates whether gap between BLS survey periods is 4 or 5 weeks in December). (iii) The great moderation period can be characterized by two negative mean shifts; one in the mid eighties, and another one in the early nineties that lasted until the early two thousands. (iv) The sub-prime crisis can be characterized by a positive mean shift in the period $2008(8) - 2010(4)$.

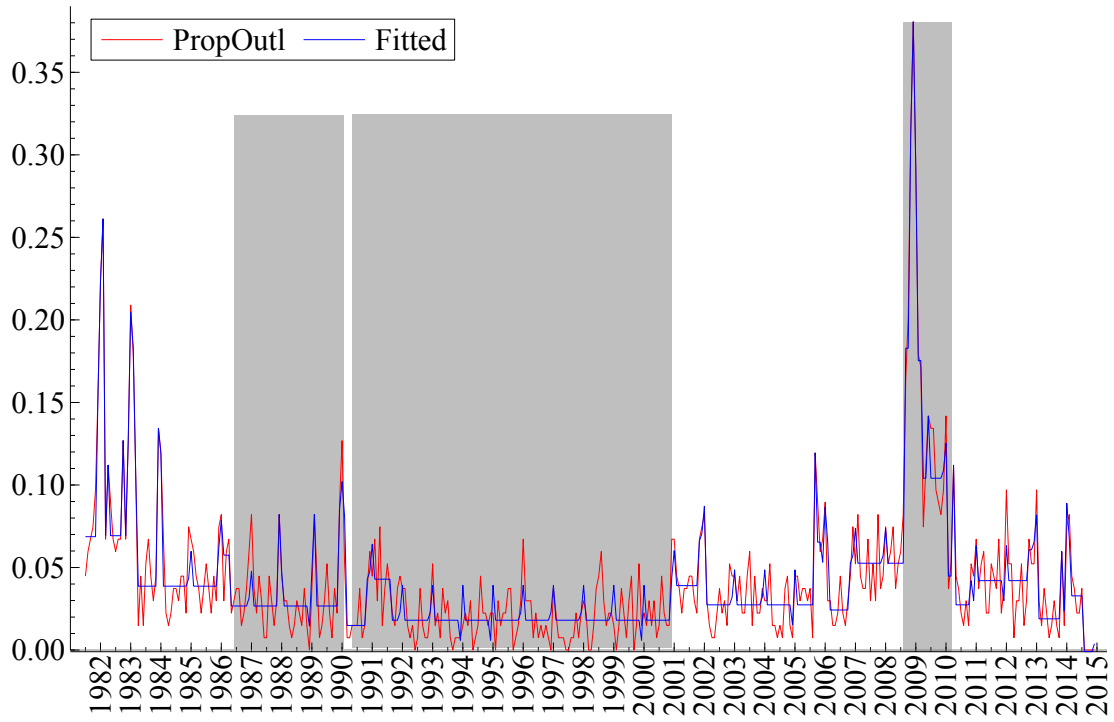



Figure VI.5: Changes in the mean proportion of series with outliers along the sample

As argued in §III.6.2, the exhaustive outliers' search we made is not only relevant for analyzing the components, but also for modeling the $IPI$ itself. This is so for outliers in the components are also outliers in the aggregate but, in many cases, they can be estimated only in the components. In order to use the individual outliers in a model for the $IPI$, we proceed as in § III.6.2, i.e., constructing the *aggregated outlier* series (*AggOut*) as the weighted sum of all individual outliers multiplied by their coefficients, and including this series in a model for the $IPI$. Next, we compare three simple models. Starting from the $GUM$; $\Delta IPI_t = c + \sum_{i=1}^{4} \phi_i \Delta IPI_{t-i} + \phi_{12} \Delta IPI_{t-12} + \sum_{c=1}^{21} \rho_i Cal_{i,t} + \epsilon_t$; where $Cal_{i,t}$ is the $i^{th}$ calender effect considered by the Board of Governors of the Federal Reserve System, we consider the following three possibilities and estimate them with *Autometrics*: (1) *Only IIS:* IIS is applied in previous GUM. (2) *Only AggOut:* the GUM is augmented with the series of $AggOut_t$ (IIS is not used). (3) *AggOut:* IIS is applied in the augmented GUM of the previous case.

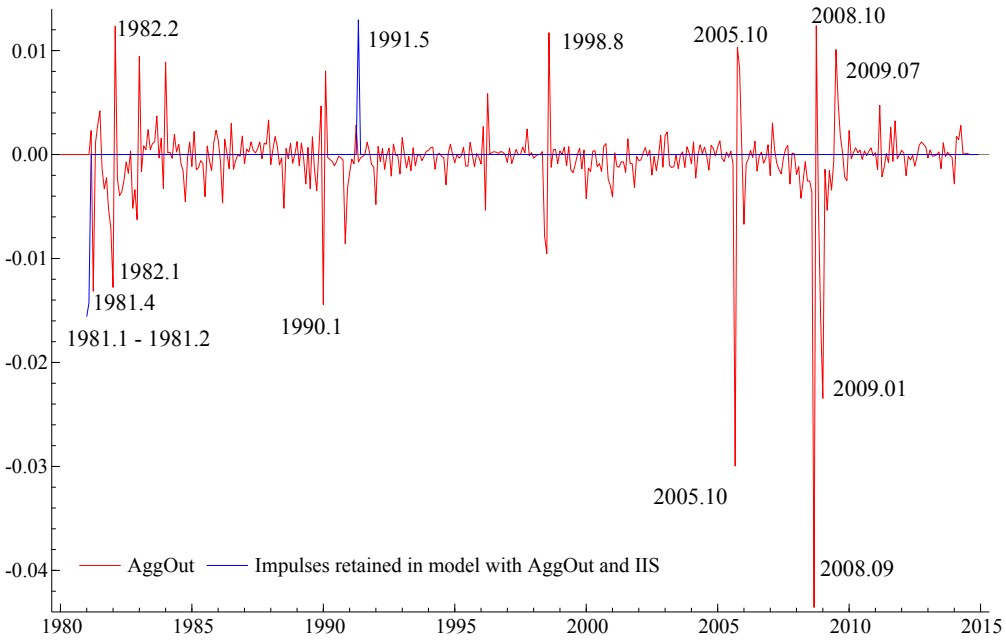
Table VI.9 includes model selection criteria for the three possibilities. As it shows, the two models including $AggOut_t$ outperform model (1). As was the case for the CPI, model (3) seems to be the best option. This last result suggest two conclusions: some components' outliers — which are also outliers of the $IPI$ — are not identifiable in the model for the aggregate, and some IPI's outliers — which must be present in some component — are not identifiable in

component's models, probably because these observations correspond to small outliers of the same sign in more than one component. Figure VI.6 includes the *aggregated outlier* series and the retained impulses of $\Delta IPI$ in model (3), blue peaks represent the outliers which are not identifiable in disaggregated series (1981.1, 1981.2, and 1991.5).

Table VI.9: Comparison of different models for the $\Delta IPI$

| | **AIC** | **SIC** | ***Adj.*$R^2$** |
|---|---|---|---|
| **Only IIS** | -7.77 | -7.62 | 0.45 |
| **Only AggOutl** | -8.28 | -8.17 | 0.65 |
| **AggOutl + IIS** | -8.39 | -8.23 | 0.70 |

Basic *GUM*: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$.
*Only IIS:* IIS is applied in previous GUM.
*Only AggOutl:* The GUM is augmented with the series of $AggOut_t$ (IIS is not used).
*AggOutl:* IIS is applied in the augmented GUM.




- *GUM*: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + AggOut_t + \epsilon_t$.
- IIS is applied in previous GUM and the resulting impulses are added up using their coefficients (blue line).

Figure VI.6: Aggregated and remaining outliers in $\Delta IPI$

### VI.5.3 Results of the pairwise tests: common trends

For the outlier's corrected series (see §III.2) we perform Johansen's tests at the 5% of significance with the number of lags for each pair determined with the *AIC* in a model with cointegration relationship and not disregarding, at the outset, a deterministic trend in the 'long run'.

Among the 134 components there are 8911 possible pairs. Without considering the three quality conditions studied in §III.4.2 (significant $\beta$ coefficients, second largest root not close to one, and stability of the cointegration relationships) we find 2250 cointegrated pairs. After disregarding those that do not satisfy any of the conditions we keep 793 pairs that are considered to satisfy stable and statistically sensible long run restrictions.

For grouping the components by blocks we consider the strategy summarized in §III.5 augmented by the *relaxation procedure* also described in that section. In order to avoid spurious grouping, we also disregard blocks with less than four series (see chapter II). In the *relaxation step* a maximum of 2 *holes* is admitted[3] and the significance level augmented to 10%.

Regarding the extensions mentioned in §III.4.1 (inclusion of a weakly exogenous variable, parsimonious long lags (PLL), and a specific form of non-linearity), we only consider the inclusion of a weakly exogenous variable. As argued in §V.4, PLL structures should not be considered for the search of common cycles. Since we want an empirical design that takes into account the possibility of common cycles from the beginning, we disregard PLL structures. On the other hand, the non-linearity considered in §III.4.1 was specific for prices, so we do not consider it in this case.

As exogenous variable we keep on using the Real Effective Exchange Rate (constructed by the Bank for International Settlements[4], using consumer price indexes for trading partners to deflate nominal effective exchange rate).

Table VI.10 summarizes the results with and without the $REER$. Since the inclusion of the $REER$ when it is not needed may distort cointegration tests, we prefer the sequential procedure in which $REER$ is included only in a second step after having constructed the blocks without considering this variable.

Table VI.11 gives some further details. To make the table legible we use the same six broad categories as in table VI.8. The main conclusion of the table is that blocks of series sharing one common unit root (or two for blocks 8 to 10 which have $REER$) cannot be assigned to a single broad category. However, in almost all the cases, more than 80% of the *Set* weight is explained by two broad categories.

This observation has two relevant implications: first the ad-hoc method proposed by Boivin and Ng (2006) for extracting non-pervasive common factors, based on pre-grouping the series

[3] If the original set has less than 6 series only one hole is admitted.
[4] https://www.bis.org/statistics/eer/index.htm

Table VI.10: Summary Pairwise procedure results: total proportion of series and weight of all the (almost) fully cointegrated subsets with outlier corrected series

Total proportion of series and weight of all the *Almost fully cointegrated sets*

| | CantSets | Num of Sers | Weight(%)* | Weight top three(%)** |
|---|---|---|---|---|
| No REER | 7 | 30 | 23.1 | 9.7 |
| With REER | 9 | 41 | 31.0 | 8.2 |
| Sequential REER[a] | 10 | 44 | 33.6 | 9.8 |

(*) Represents the proportion of weight in the 134 series considered.
(**) Weight of the tree series with larger weights.
*(a)* In this strategy blocks are firstly formed without *REER* and then the procedure with *REER* is executed only for the series not included in any previous block.

in broad categories, would not work for the US IPI; second, although a 'labeling' strategy that matches blocks with single broad categories is not possible, this could be done using just two categories.

Table VI.11: Detailed results of the Pairwise procedure with common trends: number of series and proportion of weight by broad categories and blocks

| | DurMan | | Oth Non.Dur Man | | Food,Bevg,Tob | | Chemicals | | Utilities | | Mining | | Tot W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q | W | Q | W | Q | W | Q | W | Q | W | Q | W | |
| TOTAL | 63 | 35.3 | 26 | 17.3 | 17 | 9.0 | 11 | 6.7 | 10 | 13.5 | 7 | 18.3 | |
| Block 1 | 3 | 9.1 | 1 | 5.3 | 0 | 0.0 | 0 | 0.0 | 1 | 85.6 | 0 | 0.0 | 2.8 |
| Block 2 | 4 | 64.5 | 1 | 35.5 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 2.1 |
| Block 3 | 2 | 38.7 | 0 | 0.0 | 0 | 0.0 | 1 | 16.0 | 1 | 45.3 | 0 | 0.0 | 2.0 |
| Block 4 | 2 | 27.6 | 1 | 12.9 | 0 | 0.0 | 0 | 0.0 | 1 | 59.5 | 0 | 0.0 | 3.6 |
| Block 5 | 1 | 62.9 | 1 | 9.9 | 1 | 17.7 | 1 | 9.5 | 0 | 0.0 | 0 | 0.0 | 0.7 |
| Block 6 | 1 | 24.2 | 2 | 9.1 | 0 | 0.0 | 1 | 66.7 | 0 | 0.0 | 0 | 0.0 | 2.2 |
| Block 7 | 1 | 0.2 | 1 | 44.4 | 1 | 24.1 | 0 | 0.0 | 0 | 0.0 | 1 | 31.4 | 9.7 |
| Block 8 | 3 | 34.4 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1 | 65.6 | 3.7 |
| Block 9 | 2 | 44.1 | 0 | 0.0 | 1 | 8.0 | 0 | 0.0 | 2 | 47.9 | 0 | 0.0 | 3.0 |
| Block 10 | 2 | 35.2 | 2 | 51.1 | 1 | 13.8 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 3.9 |
| TOTAL | 21 | | 9 | | 4 | | 3 | | 5 | | 2 | | 33.6 |

Columns $Q$ indicate the amount of series in each category and Set.
Columns $W$ indicate the the total weight of each category in the CPI and the proportion of the weight of each category in each Set.
Last column contains the total weight of the blocks.

### VI.5.4 Results of the pairwise tests: common cycles

Among the 134 components there are 8911 possible pairs. Without considering the aforementioned two conditions (non zero coefficients and stability) we find 1766 pairs with a common cycle. After disregarding those that do not satisfy the the two conditions conditions we keep

641 pairs that are considered to satisfy stable and statistically sensible restrictions.

Table VI.12 summarizes the results of the pairwise procedure with common cycles. It contains the amount and weight of all the series included in some 'single-cycle' subset.

Table VI.12: Summary Pairwise procedure results with common cycles: total proportion of series and weight of all the blocks (with all CIrels, all controls and *sequential* $REER^{(a)}$)

Total proportion of series and weight of all the *Almost fully connected sets*

| Num of sets | Num of Sers | | Weight(%)* | | Weight top three(%)** |
|---|---|---|---|---|---|
| | CC | $CC\&CT^{(b)}$ | CC | $CC\&CT$ | |
| 4 | 20 | 9 | 16.8 | 9.6 | 8.6 |

*(a)* In this strategy blocks are firstly formed without *REER* and then the procedure with *REER* is executed only for the series not included in any block.
*(b)* This column contains the number of series that belong to some almost fully connected set *and* some almost fully cointegrated set.
(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the IPI.
(**) Weight of the tree heavier series.

Table VI.13 gives some further details. Conclusions are similar to those from table VI.11. Blocks of series sharing one common cycle (or two for groups with *REER*) cannot be assigned to a single broad category. However, in almost all the cases, more than 85% of the *Set* weight is explained by two broad categories. The exceptions are blocks 3 to 6 for which the two most important categories explain more than 65% of the block's weights.

Table VI.13: Detailed results of the Pairwise procedure: number of series and proportion of weight by broad categories and blocks

| | DurMan | | Oth Non.Dur Man | | Food,Bevg, Tob | | Chemicals | | Utilities | | Mining | | Tot W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q | W | Q | W | Q | W | Q | W | Q | W | Q | W | |
| TOTAL | 63 | 35.3 | 26 | 17.3 | 17 | 9.0 | 11 | 6.7 | 10 | 13.5 | 7 | 18.3 | |
| Block 1 | 4 | 20.3 | 0 | 0.0 | 1 | 13.0 | 2 | 56.6 | 1 | 10.1 | 0 | 0.0 | 4.1 |
| Block 2 | 2 | 23.6 | 1 | 57.4 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1 | 19.0 | 7.5 |
| Block 3 | 1 | 4.7 | 1 | 9.6 | 1 | 8.2 | 0 | 0.0 | 0 | 0.0 | 1 | 77.4 | 3.1 |
| Block 4 | 1 | 28.3 | 2 | 33.1 | 0 | 0.0 | 1 | 38.5 | 0 | 0.0 | 0 | 0.0 | 2.0 |
| TOTAL | 8 | | 4 | | 2 | | 3 | | 1 | | 2 | | 16.8 |

Columns *Q* indicate the amount of series in each category and Set.
Columns *W* indicate the the total weight of each category in the iPI and the proportion of the weight of each category in each Set.
Last column contains the total weight of the blocks.

### VI.5.5 Forecasts for all the US IPI components with the I-PW approach

Figure VI.7, figure VI.8, table VI.14 and table VI.15 summarize the detailed analysis of the components.

Figure VI.7 includes all the components' forecasts in a single plot, and figure VI.8 includes the box plots of the 164 components for each of the 24 months of years 2013 and 2014. From these two figures, similar observations of for the case of the CPI can be done: (1) The dispersion of the forecasts of the different components seems to be smaller than the observed values of 2013, compare the sizes of the boxes and the length of the whiskers of figure VI.2 in 2013 and 2014. This is true for the bulk of forecasts closer to the headline forecasts and also for the rest which shows important calendar effects. (2) In 2013 the headline inflation seems to be around the medians of the components forecasts, but in 2014 there is a greater proportion of components forecasts below the headline forecasts, meaning that components with greater weights in the IPI are forecast with greater rates of growth.

Table VI.14 gives disaggregated forecasts for the average annual rates of growth of US IPI and its components for 2014 made with information up to December 2013. The table uses green shadows to indicate that the point forecasts of the components are above the upper bound of the confidence interval for IPI and red for the components' forecasts below the corresponding upper bound. Italics and bold letters are used to indicate that the weight of a particular component is relatively high. (see the *Notes to table VI.14* below for a detailed description).

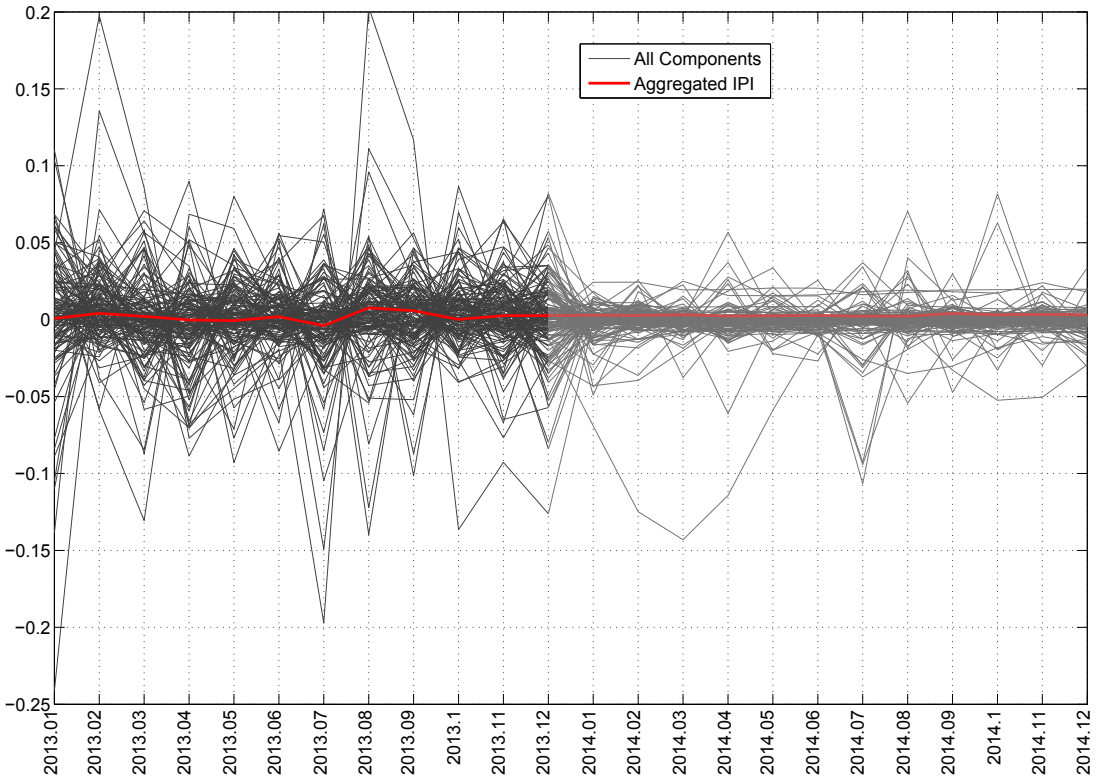
Columns $09 - 13$ contain the average growth rate for the period 2009-2013. Except for some exceptions, those columns are green or white what implies larger forecast growth rate than in previous five years, or stability, in almost all sectors. Half of the sectors in the *utilities* sub-group are not only forecast to grow above its last five years average but also above the aggregated IPI. For their weights and expected growth rates, *Crude Oil* and *Semiconductor and other electronic component*, have a special positive impact in the forecast for the IPI.

Sectors that are forecast to grow below the aggregated IPI cannot be assigned to specific sub-groups. For their weights and expected growth rates, *Aerospace product and parts* and *Tobacco*, have a special negative the the forecast for the IPI.

Table VI.15 classifies all the components according to our ability to forecast them. Red colors are used for components' which are relatively hard to forecast (large Root Mean Square Forecast Errors) and green for relatively easy to forecast ones. Additionally, highly dynamic components

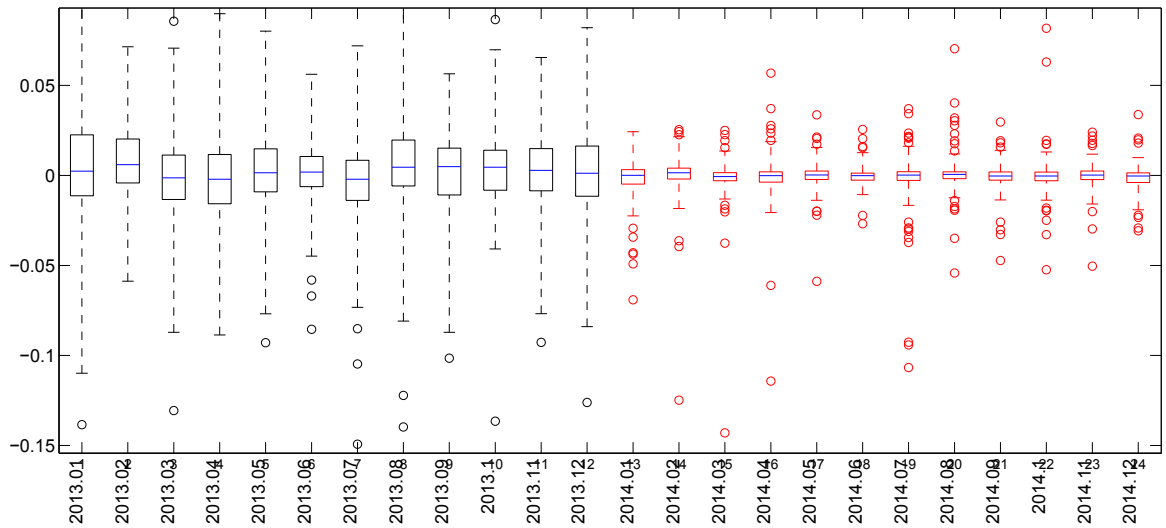
(those with green color in column *2014* of table VI.14) are marked with a dotted shadow (see the *Notes to table VI.6* above for a detailed description).

Two main conclusions can be extracted form table VI.15: (i) Food, Beverages and Tobacco products; and Minerals are, in general, relatively easy to forecast; (ii) relatively hard to forecast sectors are, in general, Durable goods.




Dark lines are observed values and light ones are forecasts.

Figure VI.7: Components' forecasts at 2013.12 ($\Delta logIPI_t$)




- The edges of the box are the $25^{th}$ ($q_1$) and $75^{th}$ ($q_3$) percentiles.
- The whiskers extend from $q_1 - 3(q_3 - q_1)$ to $q_3 + 3(q_3 - q_1)$.
- Black color indicates observed values, red, forecasts.

Figure VI.8: Box plots of the observed and forecast components at 2013.12 ($\Delta logIPI_t$)

**Notes to table VI.14**

- Columns *09.13* include the mean of annual growth of each log industrial production of years 2009 to 2013.
- Columns *2014* include the forecast of $mean(logC_{2014_{1:12}}) - mean(logC_{2013_{1:12}})$ with information up to December 2013, where $C$ stands for *component*.
- Names Reference:
  - Man Dur: Non energy industrial durable goods.
  - FBT: Food beverages and tobacco
  - Min: Mining
  - Chem: Chemicals
  - Util: Utilities
  - Oth: Other non-durable industrial goods
- Color references for columns *09-13*
  - Green: the forecast at December 2013 for 2014 is larger than the upper bound of the 80% confidence interval for the mean growth of the component in years 2009 to 2013.
  - Red: the forecast at December 2013 for 2014 is smaller than the lower bound of the 80% confidence interval for the mean growth of the component in years 2009 to 2013.
  - White: the forecast at December 2013 for 2014 is inside the 80% confidence interval for the mean growth of the component in years 2009 to 2013.
  - Standard deviation for the mean of $(\Delta_{12} logC_t)$ at December of years 2009 to 2013 is computed as $\frac{\sigma}{\sqrt{6}}$ where $\sigma$ is the sample standard deviation $\Delta_{12} logC_t$. Autocovariances at lags 12 and larger are ignored, and the confidence interval is constructed assuming normality.
- Color references for columns *2014*
  - Green: the forecast at December 2013 for 2014 is larger than the upper bound of the 80% confidence interval for the mean growth of the aggregated IPI in years 2009 to 2013.

- Red: the forecast at December 2013 for 2014 is smaller than the lower bound of the 80% confidence interval for the mean growth of the aggregated IPI in years 2009 to 2013.
- White: the forecast at December 2013 for 2014 is inside the 80% confidence interval for the mean growth of the aggregated IPI in years 2009 to 2013.
- Blue italics: indicate component's weights larger than the average weight $(1/N)$
- Blue bold: indicate component's weights larger than the $2/N$.
- Standard deviation for the 12 step ahead forecast error of $\Delta_{12} logIPI_t$ is computed as the historical out of sample Root Mean Squared Forecast Error (see table VI.7).
- Grey shadows in components' names are just to distinguish between categories.

Table VI.14: Components' forecasts ($\hat{\Delta}_{12}\log(IPI_t)$)

## IPI and all components (h=12) 2013 and 2014

| DurMan | 09-13 | 2014 | D_M &Oth | 09-13 | 2014 | FBT/Min | 09-13 | 2014 | Chem/Util | 09-13 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Miscellane | -1.2 | **3.9** | Nav. meas. | 0.2 | *0.8* | Animal foo | 0.7 | 6.3 | Org. chem. | -2.1 | *1.9* |
| Comp. & pe | -3.5 | -46.5 | Small elec | -4.2 | 7.9 | Grain & oi | -0.1 | 2.4 | Ind. gas | -8.5 | -2.6 |
| Communic. | -0.6 | 6.7 | Majorapp. | -6.9 | -2.6 | Sugar & co | -1.3 | 5.4 | Synth. dye | -2.3 | 2.1 |
| Semic. & o | 11.2 | *15.6* | Elec. ligh | -5.3 | -3.0 | Fruit & ve | -0.5 | *1.9* | Oth. inorg | -5.7 | -5.6 |
| Motor vehi | 1.3 | *4.4* | Elec. eq. | -4.2 | -3.9 | Fluid milk | -0.3 | -0.6 | Plast. mat | -2.7 | *2.8* |
| Sawmi. & w | -1.5 | 0.5 | Battery | -0.4 | 0.6 | Creameryb | 5.3 | -6.3 | Synth. rub | -4.7 | 0.6 |
| Ven.plyw. | -5.2 | 4.7 | Comm. & en | -2.7 | 1.2 | Cheese | 1.4 | 2.9 | Artif. & s | -2.6 | 6.9 |
| Millwork | -7.5 | 2.3 | Oth. elec. | -1.0 | 1.8 | Dry,cond. | 7.5 | 3.1 | Pest. fer. | 1.9 | 2.4 |
| Woodconta | -3.1 | 3.8 | Automobile | 3.4 | *3.6* | Ice cr. & | -6.8 | -8.5 | Pharm. & m | -4.5 | **-1.0** |
| Oth. wood | -6.9 | -5.0 | Light truc | -2.7 | *9.1* | Beef | -1.7 | 0.5 | Pain. & ot | -3.5 | *3.8* |
| Pott. cer. | -3.6 | -0.4 | Heavy duty | 7.2 | 4.1 | Pork | -1.3 | 3.8 | Soap cle. | -3.2 | *1.3* |
| Clay bu. m | -5.0 | 0.9 | Motor veh. | -4.5 | 7.4 | Miscell. m | -3.0 | -3.4 | Hydroel. p | 4.3 | -4.0 |
| Oth. nonme | -3.9 | 0.7 | Aeros. prd | 0.3 | **-3.6** | Poult. pro | -1.0 | *3.4* | Fossil fue | -1.0 | *3.8* |
| Lime & gyp | -6.0 | 8.1 | Railr. rol | 4.9 | 1.0 | Baker. & t | -0.6 | *0.5* | Nucelec. | -1.1 | *1.9* |
| Glass & pr | -1.2 | -2.3 | Ship &boa | -0.1 | 3.5 | Oth. food | 0.7 | *4.5* | Commer. & | 0.7 | *3.6* |
| Cement | -8.2 | 1.4 | Oth. trans | -11.5 | 3.4 | Bev. | 3.2 | *-0.2* | Ind. elec. | -1.3 | *-2.2* |
| Concrete & | -8.0 | -1.8 | House. & i | -9.8 | 6.1 | Tob. | -4.5 | *-6.3* | Resid. ele | -0.4 | *4.7* |
| Pigiron | -5.0 | 3.8 | Office& o | -3.3 | -0.7 | Crude oil | 6.5 | **4.4** | Commer. & | -0.2 | 9.4 |
| Rawsteel | -3.3 | -4.1 | Print.&rel | -5.5 | *1.6* | Natural ga | 3.8 | *0.1* | Ind. gas s | 2.1 | 6.5 |
| Coke & prd | -0.8 | -16.5 | Fib. yarn | -8.3 | -0.8 | Natural ga | 4.8 | 2.2 | Resid. gas | -1.5 | 9.5 |
| Constr. st | -5.1 | -2.2 | Fab. mills | -5.4 | -3.4 | Coal minin | -3.4 | *-0.5* | Gas transm | 0.6 | 4.8 |
| Cons. dur. | 4.1 | 10.4 | Tex. & fab | -6.4 | 0.2 | Metal ore | 1.1 | -2.2 | | | |
| Can& clo. | -8.6 | -5.5 | Tex. furn. | -9.4 | -2.0 | Nonmet. mi | -6.4 | *1.9* | | | |
| Equ. steel | -1.8 | 6.7 | Oth. tex. | -3.2 | 4.6 | Supp. act. | -1.8 | *0.4* | | | |
| Misc. stee | -4.0 | 0.6 | Apparel | -12.7 | -4.8 | | | | | | |
| Alum. & al | 0.6 | 1.3 | Leath. & a | -4.3 | -1.9 | | | | | | |
| Nonfe. met | 2.3 | -3.0 | Pulpmills | -2.9 | 0.8 | | | | | | |
| Foundries | -2.4 | -0.1 | Paper mill | -3.7 | -1.7 | | | | | | |
| Forg. & st | -0.8 | 1.4 | Paperboard | -0.4 | -0.9 | | | | | | |
| Cutl. & ha | -3.0 | -0.9 | Paperboard | -2.8 | -0.3 | | | | | | |
| Archit. & | -5.8 | *-0.3* | Pap. bag & | -2.5 | -6.2 | | | | | | |
| Hardware | -8.3 | -1.2 | Oth. conv | -1.1 | -3.5 | | | | | | |
| Spring & | -5.0 | 2.1 | Av. fuel & | -1.6 | 4.2 | | | | | | |
| Mach. shp. | -1.3 | *3.7* | Dist. fuel | 1.2 | 3.0 | | | | | | |
| Coat. eng. | -2.6 | -4.9 | Aut. gaso | 0.9 | *1.9* | | | | | | |
| Oth. fab. | -3.5 | *1.6* | Res. fuel | -10.8 | -5.4 | | | | | | |
| Agr. imp | 8.6 | -19.4 | Oth. ref. | -3.8 | 0.4 | | | | | | |
| Const. mac | -2.2 | -12.4 | Pavi. roof | -0.5 | -7.2 | | | | | | |
| Min. & oil | 3.2 | -1.3 | Plas.prd. | -1.8 | *1.9* | | | | | | |
| Ind. mach. | -6.1 | 3.3 | Tire | -3.0 | -3.4 | | | | | | |
| Comm. & se | -0.4 | *0.9* | Rubb. prd. | -2.1 | 1.3 | | | | | | |
| Vent. heat | -0.5 | -0.4 | Logging | -2.8 | 4.5 | | | | | | |
| Metalwork. | -1.2 | -2.3 | Newsp. pub | -9.3 | -4.3 | | | | | | |
| Eng. tur. | 0.8 | 1.0 | Perio. boo | -5.5 | *-2.9* | | | | | | |
| Aud. & vid | -13.9 | -5.9 | | | | | | | | | |

| | IPI | Fore(h12) |
|---|---|---|
| **2013** | **1.9** | **2.7** |
| **2014** | **3.6** | **1.3** |

| Fore Sub-Groups | 2014 |
|---|---|
| DurMan | 1.1 |
| Other Non.Dur Man | -0.2 |
| Food,Bevg and Tob | 1.5 |
| Mining | 2.5 |
| Chemicals | 0.6 |
| Utilities | 3.6 |

**confidence 80%**

2014

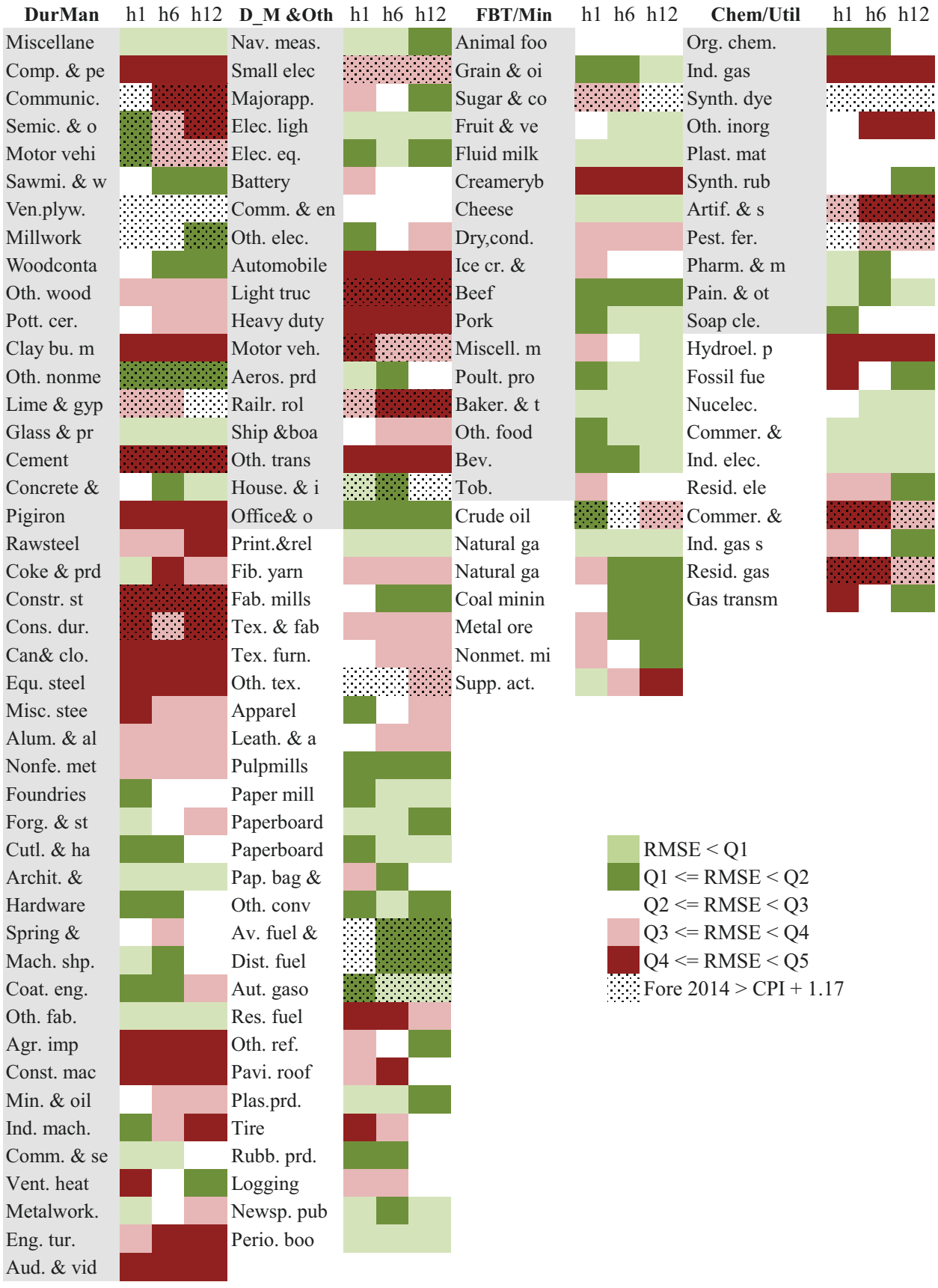
- Fore > IPI - 2.7
- Fore = IPI +- 2.7
- Fore < IPI + 2.7
- *italics* Weight > 100/134
- **bold** Weight larger than 3%

**confidence 80%**

2008-2013

- Fore 14 > Mean(08-13)
- Fore 14 = Mean(08-13)
- Fore 14 < Mean(08-13)

Table VI.15: Components RMSFE



### VI.5.6 Comparative results

For a complete description of the forecasting exercises see §VI.3. Table VI.16 includes a summary of the forecasting results. The first row includes the root mean squared forecast error in the direct baseline for $\Delta_{12}log(IPI)$ for horizons $H = 1$ to $H = 12$. All the other entries in the table are ratios with respect to the baseline. The evaluation period is $2006.1 - 2014.12$.

As illustrated in figure VI.5, during years 2008 and 2009 there was significant increase in the number of series with outliers. Thus, in table VI.17 we repeat table VI.16 but excluding those years. From the comparison of the two tables three main conclusions emerge: (i) On average, the RMSFE in the baseline are 55% smaller when 2008 and 2009 are disregarded. (ii) This reduction is even more relevant for indirect approaches, and for the pairwise procedure in particular. For the case of *I-CC&CT-IPI*, RMSFE are on average 67% smaller when 2008 and 2009 years are not considered. (iii) This observations reinforce the conclusion that the problem of outliers is more important when dealing with disaggregates. When forecasting with the indirect approach it could be very important to apply the robustifying procedure initially proposed by Hendry (2006) and studied in Castle et al. (2015).

In what follows we will focus on the results of table VI.17.

#### Direct approaches

*Disaggregated information (D-DI-1 and D-DI-2)*

When using disaggregated information in a scalar model for the aggregate, as proposed by Hendry and Hubrich (2011), in contrast to the case of the CPI, high disaggregation levels are preferred to lower ones. Nonetheless the use of disaggregated information does not improve the results of the baseline.

*Scalar models enlarged with Dynamic Factors (D-DFM-1)*

The best results are obtained with three factors. The inclusion of the factor produces some gains over the baseline in horizons 1 to 8. In larger horizons the baseline is preferred.

#### indirect approaches

*Univariate models for all the components enlarged with the aggregated IPI (I-IPI)*

In short horizons it produces some improvements with respect to the baseline, particularly

in horizons 4 and 5, for which RMSFE are 15% and 10% smaller. In long horizons (9-12) these improvements disappear; the baseline does better.

*Pairwise procedures*

In these procedures the basic components that do not belong to a subset sharing one common trend, one common cycle or both, are forecast using univariate individual models. An alternative consists in forecasting just the aggregate of those basic components ($I - PW - GP$). In short horizons the latter alternative provides slightly better forecast for the aggregated IPI than the corresponding case in which basic components outside subsets with common features are forecast individually. For longer horizons the differences in favor of *I-PW-GP* become larger.

When considering only common unit roots restrictions results are better than in the basic indirect procedure (I) in almost all horizons. This dominance is particularly relevant for horizons 6 to 12 for which differences in RMSFE reach a maximum around 8%[5]. Additionally, when considering only common unit roots, indirect procedures outperform direct procedures in all horizons (except for *D-DFM-3* at $H = 1$).

If instead of common unit roots we consider only common cycles, previous conclusions are still valid.

Overall, *I-CC& CT-GP* seems to be the clear winner; it dominates all the other procedures in all horizons, except for $H = 2$ and $H = 4$ in which *I-onlyCT-IPI* is somewhat better. Though dominated by *I-CC& CT-GP*, it is important to note that *I-CC& CT* also seems to dominate the other possibilities.

The came conclusion as for the case of the CPI is also valid now. Modeling of the basic components by single-equation methods taking into account the common features restrictions between them, identified by pairwise methods, can be considered as adequate in the sense that the forecasts for the aggregate are quite good. This is an interesting indirect test of the common-feature disaggregated approach.

[5]As the figures are ratios with respect to the baseline, this 8% refers to the RMSFE of the baseline

Table VI.16: Relative RMSE $\Delta_{12}log(IPI)$. Evaluation period: 2006.1 - 2014.12 (first row: RMSE for the baseline. All the others are ratios with respect to the first)

| | H=1 | H=2 | H=3 | H=4 | H=5 | H=6 | H=7 | H=8 | H=9 | H=10 | H=11 | H=12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | **0.73** | **1.08** | **1.43** | **1.86** | **2.36** | **2.86** | **3.38** | **3.88** | **4.39** | **4.91** | **5.42** | **5.90** |
| D-DI-1 | 0.94 | 1.00 | 1.03 | 1.05 | 1.07 | 1.11 | 1.11 | 1.11 | 1.11 | 1.11 | 1.10 | 1.10 |
| D-DI-2 | 0.85 | 0.89 | 1.08 | 1.03 | 1.04 | 1.09 | 1.09 | 1.09 | 1.09 | 1.08 | 1.07 | 1.06 |
| D-REER | 1.00 | 1.01 | 1.02 | 1.04 | 1.03 | 1.04 | 1.04 | 1.04 | 1.03 | 1.03 | 1.03 | 1.02 |
| D-DFM-3 | 0.88 | 0.92 | 0.97 | 1.00 | 1.03 | 1.08 | 1.13 | 1.18 | 1.22 | 1.25 | 1.28 | 1.31 |
| I-IPI | 0.98 | 1.01 | 1.08 | 1.11 | 1.12 | 1.14 | 1.14 | 1.14 | 1.14 | 1.14 | 1.13 | 1.13 |
| **Pairwise** | | | | | | | | | | | | |
| I-CC&CT-IPI | 1.25 | 1.25 | 1.35 | 1.28 | 1.25 | 1.24 | 1.23 | 1.21 | 1.20 | 1.18 | 1.17 | 1.15 |
| I-CC&CT-IPI-GP | 1.25 | 1.21 | 1.37 | 1.27 | 1.24 | 1.24 | 1.22 | 1.20 | 1.19 | 1.17 | 1.15 | 1.13 |
| I-onlyCT-IPI | 0.99 | 1.02 | 1.09 | 1.12 | 1.13 | 1.14 | 1.14 | 1.14 | 1.14 | 1.14 | 1.13 | 1.13 |
| I-onlyCT-IPI-GP | 0.97 | 0.99 | 1.02 | 1.05 | 1.07 | 1.08 | 1.08 | 1.09 | 1.09 | 1.08 | 1.08 | 1.07 |
| I-onlyCC-IPI | 0.98 | 1.04 | 1.11 | 1.14 | 1.14 | 1.15 | 1.15 | 1.15 | 1.14 | 1.13 | 1.13 | 1.12 |
| I-onlyCC-IPI-GP | 0.97 | 1.01 | 1.05 | 1.08 | 1.09 | 1.09 | 1.09 | 1.10 | 1.09 | 1.09 | 1.08 | 1.07 |

- See tables VI.1 and VI.2 for a description of each exercise.
- Dark red entrances highlight the loser procedure.
- Light red indicates procedures that are, at most, 5 points smaller than the worst one.
- Dark green indicates the best procedure.
- Light green indicates procedures that are, at most, 5 points larger than the best one.

Table VI.17: Relative RMSE $\Delta_{12}log(IPI)$. Evaluation period: 2006.1 - 2014.12, excluding years 2008 and 2009 (first row: RMSE for the baseline. All the others are ratios with respect to the first)

| | H=1 | H=2 | H=3 | H=4 | H=5 | H=6 | H=7 | H=8 | H=9 | H=10 | H=11 | H=12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | **0.42** | **0.55** | **0.69** | **0.84** | **0.99** | **1.12** | **1.41** | **1.71** | **1.92** | **2.10** | **2.27** | **2.42** |
| D-DI-1 | 1.01 | 1.07 | 1.11 | 1.10 | 1.14 | 1.25 | 1.10 | 1.09 | 1.12 | 1.17 | 1.22 | 1.56 |
| D-DI-2 | 1.02 | 1.03 | 1.10 | 1.07 | 1.14 | 1.12 | 1.07 | 1.05 | 1.07 | 1.10 | 1.14 | 1.28 |
| D-REER | 1.02 | 1.02 | 0.98 | 0.98 | 0.99 | 1.00 | 1.00 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 |
| D-DFM-3 | 0.95 | 0.96 | 0.98 | 0.90 | 0.97 | 0.95 | 0.91 | 0.99 | 1.08 | 1.25 | 1.33 | 1.85 |
| I-IPI | 0.97 | 0.91 | 0.91 | 0.85 | 0.90 | 0.97 | 0.97 | 0.97 | 1.01 | 1.03 | 1.05 | 1.05 |
| **Pairwise** | | | | | | | | | | | | |
| I-CC&CT-IPI | 0.96 | 0.91 | 0.93 | 0.87 | 0.91 | 0.96 | 0.93 | 0.89 | 0.88 | 0.89 | 0.89 | 0.87 |
| I-CC&CT-IPI-GP | 0.95 | 0.92 | 0.89 | 0.89 | 0.89 | 0.93 | 0.87 | 0.83 | 0.81 | 0.83 | 0.82 | 0.81 |
| I-onlyCT-IPI | 0.96 | 0.91 | 0.91 | 0.84 | 0.90 | 0.94 | 0.93 | 0.91 | 0.94 | 0.96 | 0.98 | 0.97 |
| I-onlyCT-IPI-GP | 0.90 | 0.92 | 0.90 | 0.92 | 0.95 | 1.08 | 1.03 | 1.06 | 1.07 | 1.15 | 1.17 | 1.17 |
| I-onlyCC-IPI | 0.98 | 0.92 | 0.94 | 0.88 | 0.93 | 0.98 | 0.96 | 0.93 | 0.93 | 0.94 | 0.94 | 0.93 |
| I-onlyCC-IPI-GP | 0.99 | 1.03 | 1.02 | 1.01 | 1.03 | 1.06 | 1.01 | 0.99 | 0.99 | 1.03 | 1.06 | 1.07 |

- See tables VI.1 and VI.2 for a description of each exercise.
- Dark red entrances highlight the loser procedure.
- Light red indicates procedures that are, at most, 5 points smaller than the worst one.
- Dark green indicates the best procedure.
- Light green indicates procedures that are, at most, 5 points larger than the best one.

## VI.6 Concluding remarks

In this chapter we extended the application of the pairwise procedure to the US CPI initiated in chapter III and provided a full application for the US IPI.

In regards to the US CPI we looked for 'single-cycle' subsets and found 15 relatively small subsets containing 92 series (out of the 164), representing 30% of the CPI's weight. We applied the pairwise approach in a forecasting exercise and compared its ability to forecast the aggregate with other direct and indirect approaches. The results show not only that the inclusion of common features' restrictions improve the forecasting accuracy of indirect approaches, but also that it helps to beat direct procedures.

The complete application for the US IPI included an analysis of outliers, a search of 'fully cointegrated' and 'single-cycle' subsets, and a forecasting exercise.

The distribution of the proportion of components with outliers by date is far from uniform and shows interesting mean shifts. The 'great moderation' period can be characterized by two negative mean shifts; one in the early eighties and another one in the early nineties that lasted until the early two thousands. The other interesting mean shift is the one associated with the sub-prime crises. In August 2008 there was a positive mean shift that lasted until April 2010.

As in the case of the CPI we also found that the *aggregated outlier* — a series constructed by aggregating the outliers of the components — helps to improve the model of the aggregate. This is because the are some outliers that are not identifiable in the aggregated series.

In relation with the pairwise cointegration tests we found 10 relatively small subsets that contain 44 components (out of the 134) that represent 33.6% of the IPI. For the case of common cycles, we found 4 subsets with 20 series and 16.8% of the weight. Using this results, we compared the ability of the pairwise approach to forecast the aggregated IPI with other indirect and direct procedures. The general conclusions were similar to those obtained for the CPI; the consideration of common features' restrictions helps to improve indirect forecasts and to beat direct procedures.

# Appendix D

# Appendix to chapter VI

## D.1 Further results for the US IPI

Table D.1: Model for the proportion of series with outliers, US IPI. 1981(2) - 2014(12)

| | Coefficient | Std.Error | t-value | t-prob | Part.$R^2$ |
|---|---|---|---|---|---|
| $PWHTIMING\_SAT\_12$ | -0.0168 | 0.006 | -2.670 | 0.008 | 0.020 |
| CSeasonal | 0.0212 | 0.003 | 7.300 | 0.000 | 0.130 |
| I:1982(2) | 0.0585 | 0.021 | 2.780 | 0.006 | 0.021 |
| I:1982(4) | 0.0426 | 0.016 | 2.710 | 0.007 | 0.020 |
| I:1982(10) | 0.0576 | 0.016 | 3.650 | 0.000 | 0.036 |
| I:1982(12) | -0.0444 | 0.019 | -2.380 | 0.018 | 0.016 |
| I:1983(12) | 0.1078 | 0.016 | 6.870 | 0.000 | 0.118 |
| I:1984(1) | 0.0596 | 0.015 | 3.910 | 0.000 | 0.041 |
| I:1987(12) | 0.0507 | 0.015 | 3.370 | 0.001 | 0.031 |
| I:1989(2) | 0.0554 | 0.015 | 3.710 | 0.000 | 0.037 |
| I:2005(9) | 0.0540 | 0.016 | 3.340 | 0.001 | 0.031 |
| I:2009(6) | 0.0377 | 0.016 | 2.430 | 0.016 | 0.016 |
| I:2013(11) | 0.0408 | 0.015 | 2.640 | 0.009 | 0.019 |
| S1:1981(11) | -0.0834 | 0.016 | -5.130 | 0.000 | 0.069 |
| S1:1981(12) | -0.0507 | 0.021 | -2.400 | 0.017 | 0.016 |
| S1:1982(2) | 0.1334 | 0.016 | 8.330 | 0.000 | 0.163 |
| S1:1982(11) | -0.1142 | 0.012 | -9.590 | 0.000 | 0.206 |
| S1:1983(2) | 0.0790 | 0.018 | 4.360 | 0.000 | 0.051 |
| S1:1983(3) | 0.0658 | 0.015 | 4.390 | 0.000 | 0.052 |
| S1:1985(11) | -0.0188 | 0.007 | -2.640 | 0.009 | 0.019 |
| S1:1986(4) | 0.0309 | 0.007 | 4.400 | 0.000 | 0.052 |
| S1:1989(11) | -0.0543 | 0.009 | -6.120 | 0.000 | 0.095 |
| S1:1990(2) | 0.0660 | 0.010 | 6.570 | 0.000 | 0.109 |
| S1:1990(10) | -0.0280 | 0.007 | -4.000 | 0.000 | 0.043 |
| S1:1991(8) | 0.0249 | 0.005 | 5.110 | 0.000 | 0.068 |
| S1:2000(11) | -0.0210 | 0.005 | -4.500 | 0.000 | 0.054 |
| S1:2001(10) | -0.0270 | 0.010 | -2.800 | 0.005 | 0.022 |
| S1:2002(1) | 0.0386 | 0.009 | 4.360 | 0.000 | 0.051 |
| S1:2005(8) | -0.0380 | 0.007 | -5.400 | 0.000 | 0.076 |
| S1:2006(2) | 0.0411 | 0.008 | 4.850 | 0.000 | 0.062 |
| S1:2006(10) | -0.0283 | 0.006 | -4.650 | 0.000 | 0.058 |
| S1:2008(8) | -0.1302 | 0.011 | -12.000 | 0.000 | 0.287 |
| S1:2008(10) | -0.1306 | 0.018 | -7.230 | 0.000 | 0.129 |
| S1:2008(11) | -0.0625 | 0.021 | -2.990 | 0.003 | 0.025 |
| S1:2008(12) | 0.1060 | 0.021 | 5.020 | 0.000 | 0.066 |
| S1:2009(1) | 0.0945 | 0.018 | 5.170 | 0.000 | 0.070 |
| S1:2009(3) | 0.0713 | 0.012 | 6.180 | 0.000 | 0.097 |
| S1:2010(1) | 0.0593 | 0.012 | 5.140 | 0.000 | 0.069 |
| S1:2010(3) | -0.0672 | 0.018 | -3.720 | 0.000 | 0.038 |
| S1:2010(4) | 0.0846 | 0.016 | 5.310 | 0.000 | 0.074 |
| S1:2010(10) | -0.0147 | 0.007 | -2.170 | 0.031 | 0.013 |
| S1:2012(9) | -0.0187 | 0.008 | -2.320 | 0.021 | 0.015 |
| S1:2013(1) | 0.0418 | 0.009 | 4.770 | 0.000 | 0.060 |
| S1:2013(12) | -0.0488 | 0.012 | -4.240 | 0.000 | 0.048 |
| S1:2014(2) | 0.0349 | 0.012 | 2.810 | 0.005 | 0.022 |
| S1:2014(7) | 0.0338 | 0.009 | 3.620 | 0.000 | 0.036 |
| Constant | 0.0008 | 0.007 | 0.125 | 0.901 | 0.000 |

- $PWHTIMING\_SAT\_12$ is a variable that indicates whether gap between BLS survey periods is 4 or 5 weeks in December.
- Steps (S1:) takes the value one from the first observation until the date indicated in the name of the step, and zero form then on.
- CSeasonal is the centered seasonal variable corresponding to January. It takes the value $1 - 1/12$ in January and $-1/12$ otherwise.

# Chapter VII

# General Conclusions

In this final chapter, we summarize the conclusions of chapters II to VI, give a comment on the limitations of the research presented in this thesis and propose some future research lines.

## VII.1 Summary of conclusions

In this thesis we developed an econometric procedure to model and forecast all the components of a macro or business aggregate with common features (trends and cycles). Our strategy consists of trying to discover blocks of components with single common features, and then include the restrictions derived from these commonalities in single-equation models for all the components.

For discovering those blocks we perform common features tests between all possible pairs of components and group the components according to the test results. As official sub-aggregates could add up series that do not share common features, this approach requires working with the components maximum level of disaggregation available (*basic components*).

Since the number of basic components of macroeconomic aggregates is usually large (hundreds) this pairwise strategy requires thousands of common features tests (the number of pairs between, say, 150 components is 11,175). This large amount of tests raises several operative and statistical problems. Among the latter, the possibility of increased false rejection probability (multiple testing problem), and the estimation of partial models are the most important. A central objective of this thesis is to study the statistical foundations of the pairwise approach, and design an operational strategy to apply it.

In chapter II, we studied the properties of the pairwise procedure for testing cointegration. This procedure allows to discover blocks of series that share a unique common trend (*fully*

*cointegrated subsets*, hereafter *FC*). The consequences of the multiple testing problem, when it exists, are different depending on the type of pairs. We need to distinguish three different cases: (i) Pairs between series inside the same FC subset. Calling $r^*$ the true cointegration rank, in these pairs $r^*$ is 1. (ii) Pairs between a series inside and a series outside a FC subset ($r^* = 0$). (iii) Pairs between outsiders ($r^* = 0$).

We perform Johansen's cointegration tests in the usual sequential fashion; testing first $r = 0$ *vs.* $r > 0$ and then, if the null is rejected, $r = 1$ *vs.* $r = 2$. For pairs in (i), asymptotic power equal to one implies that finding $r = 0$ is not an issue. The problem could be wrongly rejecting $r = 1$ in favor of $r = 2$. The main theoretical result is that cointegration tests for these pairs are asymptotically equivalent, in the sense that the probability that all tests deliver the same conclusion is close to 1 as $T$ goes to infinity, independently of the number of series. Thus, multiple testing is not an issue for pairs of components inside a FC subset.

For pairs in (ii) the multiple testing problem could lead us to include wrong series in the estimated FC subsets. We showed that although there could be a multiple testing problem in this case, the expected number of wrong series included estimated FC subset is bounded by $(N-n_1)\times\varphi$, where $N$, $n_1$ and $\varphi$ are total number of basic components, the size of the FC subset and the nominal size used in the cointegration tests, respectively. Thus for, say, $N - n_1 = 100$ and $\varphi = 0.01$, we will include, at most, just one wrong series in the FC subset.

For pairs in (iii) the multiple testing problem could lead to find FC subsets formed by wrong series. We showed however that this issue can be easily avoided by choosing tight significance levels and disregarding small FC subsets.

In a Monte Carlo experiment, we confirmed the asymptotic results and compared the performance of the pairwise approach for discovering FC subsets with a DFM alternative. This comparison showed that the pairwise procedure dominates in situations in which the size of the FC subsets is relatively small with respect to the total number of components. The DFM alternative fails in those situations. For relatively large FC blocks and $T$, the DFM alternative may be preferred, though the pairwise approach also performs very well in those scenarios.

The second statistical problem associated with the pairwise strategy is the estimation of partial models. We assume the existence of a (big) VAR model for all the components, but test for cointegration in several partial bi-variate models. In a small Monte Carlo experiment we showed that when cointegration relationships are pairwise detectable, the pairwise strategy

leads to power improvements with respect to a regular Johansen's test applied to a (reduced) group of series that share a common trend in short samples.

Other relevant results in chapter II are those related to the specification of the bivariate sub-models in which cointegration tests are performed. First, we argued that changing the normalization of the cointegrating matrix may lead to an improvement in the forecasting accuracy of the individual models for the components. This potential improvement is derived from a possible reduction in the number of regressors — which does not cause information losses — and, hence, in the estimation uncertainty of those models. Second, we established the conditions under which the forecasting strategy in Espasa and Mayo-Burgos (2013) (who, in the single-equation models, include just one cointegration restriction with respect to the sub-aggregate of the fully cointegrated components) is correct. As the conditions are quite stringent, we proposed a generalization of that strategy that does not rely on them. Our proposal is to consider all the cointegration relationships as potential regressors for the forecasting equations of fully cointegrated components and to select the significant ones by applying the model selection algorithm *Autometrics*.

In chapter III we proposed a robustification strategy for the pairwise approach to discover common trends proposed in chapter II, and applied it in a real data exercise. The robustification includes a strategy for dealing with data irregularities and with short samples issues as well as some extensions to the design of the pairwise cointegration tests.

If dealing with outliers and breaks is a crucial issue of empirical analysis, it is even more important when working with disaggregates, as they tend to be more 'contaminated' than the aggregates. This issue is specially troublesome for cointegration analysis as the inclusion of step dummies in the model affect the asymptotic distribution of the tests, and the effects depend both on the number and the position of the breaks. New critical values must be simulated for each particular situation. Thus, given the large amount of tests to be performed, the traditional way for dealing with location shifts in cointegration is unfeasible in the pairwise approach.

For tackling this problem we designed a procedure that combines the saturation techniques for finding outliers and breaks proposed by Santos et al. (2008) with the GLS procedure for testing cointegration proposed by Lütkepohl et al. (2004). In a first step we make an exhaustive search of outliers in all the components, and then we use the outliers' location in the GLS procedure. This strategy does not require simulating new critical values. We showed its good

performance by Monte Carlo.

In the applications to the US CPI and the US Industrial production Index, we showed that the exhaustive search of outliers is also relevant even for modeling just the aggregate. This is so because some outliers are detectable only in the components. Additionally, this exhaustive search allows to study changes in the cross sectional distribution of the outliers along time. We did not provide an important progress in this respect. We just studied the changes in the proportion of components with outliers along time and, apart from a seasonal pattern, we found a negative mean shift during the great moderation and a positive one in the recent sub-prime crises period.

Another problem of the pairwise approach when applied to real data is how to deal with short samples. We found a relevant potency deterioration in those situations. In chapter III we proposed a strategy that mitigates this problem with small costs in terms of gauge. The proposal consists of letting enter the estimated FC subsets all the series that showed a cointegration relationship with *almost* all the other series in the subset.

In chapter III we also proposed a strategy for extending the pairwise cointegration tests with the inclusion of a weakly exogenous variable. We showed that FC subsets can still be constructed with the pairwise approach. Finally, in this chapter we considered the possibility of parsimonious long lag structures as suggested by Aron and Muellbauer (2013), and found that they can be useful in real data applications.

In chapter IV we provided a review on common cyclical features, analyzed the relationships between VAR models with common features restrictions and Dynamic Factor Models (DFM), proposed a strategy for dealing with common cycles in the framework of the pairwise approach, studied the properties of this strategy, and compared its performance with a DFM alternative.

Although the only definition of common cyclical feature that is equivalent to common cycles in the Beverdige-Nelson sense is the original *serial correlation common feature* (SCCF), proposed by Engle and Kozicki (1993), we used the term 'common cycle' in a broad sense to refer to any kind of common cyclical feature.

Our strategy for dealing with common cyclical features in the framework of the pairwise approach is similar to that for dealing with common trends; test for common cyclical features between all the possible $N(N-1)/2$ pairs that exist in a set of $N$ variables and look for subsets in which all pairs showed a common cycle (*single-cycle subsets, hereafter SC*).

For this procedure to be sensible, the definition of common cycle has to be transitive. We need that in a group of three variables, the fact that two pairs of series — out of the possible three — have a common cycle, implies that the third pair also has it. We showed that, although the only transitive definition in the context of the pairwise procedure is the original SCCF, the other definitions can also be transitive after a simple extension of the testing strategy. For the case of *weak serial correlation common features* (WF), this extension consists of including in each pairwise test not only the cointegration relationship of the pair (if it exist) but also all the other cointegration relationships of the whole system that are relevant for the series of the pair. This strategy ensures transitivity, but raises some statistical and operational problems. In §IV.5.2 we studied those problems by Monte Carlo and concluded that the cointegration relationships to be included in each pairwise test should be selected in a first step using some model selection procedure, for example, *Autometrics*.

Once the lack of transitivity is solved, the statistical and operational problems that appeared for the case of cointegration also appear in the case of common cycles. The most important statistical issues are; first, given the large amount of tests we are performing, we may be inflating the false rejection probability — multiple testing problem. Second, what are the costs of proceeding in a pairwise fashion in comparison with a full model strategy, when the number of variables is reduced and the latter procedure feasible.

Again, we studied the multiple testing issue for three different types of pairs: (i) Pairs between two series inside the same SC subset. Calling $s^*$ the true number of common cycles restrictions, for these pairs $s^*$ is 1. (ii) Pairs between a series inside and a series outside a SC subset ($s^* = 0$). (iii) Pairs between two outsiders ($s^* = 0$).

For performing the common cycles test we proceed in the traditional sequential fashion, testing first $s > 0$ *vs.* $s = 0$ and then, if the null is not rejected, $s = 2$ *vs.* $s = 1$. For the pairs in (i) the true hypothesis is $s = 1$, so that the multiple testing problem may lead to over reject $s > 0$ in favor of $s = 0$. For the other two types of pairs, the true hypothesis is $s = 0$, so that, asymptotically, there cannot be a multiple testing issue, for asymptotic power is equal to one. Therefore, the risk of including wrong series in estimated SC subsets is, asymptotically, not relevant.

In contrast with our results in chapter II, we showed that the multiple testing problem is in indeed present inside SC subsets. The reason is that the fact that the series inside those subsets

share the 'cycle', does not ensure that the pairwise test statistics are going to be dependent between each other. On the contrary, we showed that, under some conditions, they can be independent. The absence of a high degree of dependency makes the probability of including a large proportion of the true series inside the estimated SC subset to be a decreasing function of the true size of the subset. This is an undesirable property, for we want our procedure to work well not only when subsets are small, but also when they are large.

We showed that a simple 'relaxation' strategy solves the problem, with almost zero cost. The relaxation is the same as that described above for cointegration, with the difference that now it is asymptotically costless because asymptotic power is equal to one.

Regarding the comparison with the full model strategy we performed a Monte Carlo study in which the DGP was a small dimensional VAR. The results showed that, when the common cycles are pairwise detectable, nothing is lost for proceeding by pairs. On the contrary, relevant power gains for discovering the true number of common cycles relationships can emerge from the pairwise procedure in short samples.

In chapter IV we also compared the performance of the pairwise procedure with a DFM alternative, for stationary and non-stationary processes, and analyzed the results under the light of the aforementioned study of the relationship between restricted VAR models and DFM. The main conclusion was that, while the DFM fail to discover the SC subsets when their sizes is not large, the pairwise approach shows a good performance in any situation.

At this point, it is worth to describe with some more detail the simulation studies performed along this thesis. In order to assess the performance of the pairwise approach to discover FC and SC subsets, and to compare it with a DFM alternative, we simulated 100-dimensional models with common trends and common cycles. For each of the 4950 pairs that exist between the 100 simulated series we need to estimate bi-variate VAR models, for which the lag-length is unknown. When looking for FC subsets we determine the lag-length with the AIC (admitting a maximum of five lags) in models where one cointegration relationship is assumed. Then, we test for cointegration using the lag length determined in previous step. If the number of Monte Carlo replications is 1000, this implies $1000 \times 4950 = 4.95$ *million* cointegration tests. Conditional on these results, we test for (weak) common cycles in all pairs (other 4.95 million contrasts). Since we considered three different sample sizes and three different scenarios the number of contrasts rises to $4.95 \times 2 \times 3 \times 3 = 89.1$ *million*. To this number of tests we need to

add the computational effort of determining the lag length in each of the 4.95 million models for each scenario and sample size.

Finally, in chapter VI we completed the application to the US CPI of chapter III with the consideration of common cycles and a forecasting exercise, and provided a complete application to the US IPI including an analysis of outliers, a search of 'fully cointegrated' and 'single-cycle' subsets, and a forecasting exercise.

In the forecasting exercises, we compared the ability of the pairwise approach to forecast the aggregate with other indirect and direct procedures. In the pairwise approach we forecast all the components with single-equation models that included the cointegration and common cycles restrictions derived from the FC and SC subsets. When the component belongs to some FC subset, we considered all the cointegration relationships of the subset as potential regressors. When the components do not have common cycles restrictions, apart from the cointegration relationships (when applicable), we also considered own lags and lags of the other components as potential regressors and selected the final forecasting models with *Autometrics*. In all cases we used Impulse Indicator Saturation for the outliers' treatment.

The general conclusions were similar in both applications; the consideration of common features' restrictions helps to improve indirect forecasts and to beat direct procedures. An interesting extension of the application, that we did not considered, could be to implement the robustification strategy proposed by Hendry (2006) and Castle et al. (2015).

## VII.2 Limitations and extensions

The most evident limitation of our procedure is that we only look for blocks of components sharing single common features. Since the existence of a single common feature among a subset of components implies, in principle, the largest number of restrictions in the dynamics of the series, it seems to be the most interesting situation. We acknowledge however that we may be neglecting restrictions that may arise between subsets of components that share more than one common feature of the same type. A generalization of the pairwise procedure that tests for common features between all possible triples or quartets would be computationally too demanding as the number of tests will be extremely large for moderately large $N$ (for $N = 100$ the number of pairs is 4950; the number of triplets 61700; and the number of quartets is 3921225). Thus, in order to overcome this limitation some sort of block search should be

designed.

Another limitation of our procedure is that the series cannot have more than one unit root. The cases of $I(2)$ or seasonally integrated series was not considered in this thesis. An extension in this line could be interesting.

The third limitation is related to the outliers' correction procedure. Lütkepohl et al. (2004) develop the asymptotic theory for the case of a single break at unknown time. Although they argue that in the case of additional breaks the same procedure could be applied, they do not develop the asymptotic theory and neither do we. Instead, we provide some Monte Carlo evidence that suggests that the procedure performs reasonably in the case of multiple breaks. The derivation of the asymptotic theory would be an interesting extension.

Restricting the focus to common trends and common cycles is also a limitation. As we showed in §III.6.2 and §VI.5.2, the time distribution of the components' level shifts is far from being uniform. This could suggest the existence of co-breaking structures (see Hendry and Massmann (2007)), but we are not considering them. Additionally, commonalities are not necessarily restricted to the mean, but could be also present in the variance of the series. In this respect the consideration common variance structures (see, e.g; Engle and Kozicki (1993)) could be also of interest.

The general design of the Monte Carlo experiments that we performed could also be seen as a limitation. A fair critique in all the cases could be that the DGPs are quite simple, and it would be interesting to see how the results change under other type of processes. A non-exhaustive list of interesting modifications of the DGPs could be: i) There are groups of components shearing more than one common feature (factor) of the same type. ii) Idiosyncratic components of the DFM could show time dependence. iii) Idiosyncratic components of the DFM could be non-stationary. iv) The covariance matrix of the VAR's residuals could be non-diagonal. v) The coefficients' matrices of the VAR could be more complex. vi) There could be some I(2) components. vii) Some roots of the VAR's characteristic polynomial could be outside, but close, to the unit circle. viii) Conditional variance of some components' stationary transformation is time varying.

As we argued above, our Monte Carlo experiments involve several million of contrasts, making the computing time an important constraint. This is the main reason for which we did not consider other possible DGPs. The consideration of some of the aforementioned possibilities

would be an interesting extension.

Finally, another relevant extension of the simulation experiments would be a forecasting comparison between our proposal and a DFM alternative.

# References


Alchian, A. A. (1969). Information costs, pricing, and resource unemployment. *Economic Inquiry 7*(2), 109–128.

Altissimo, F., B. Mojon, and P. Zaffaroni (2007). Fast micro and slow macro: Can aggregation explain the persistence of inflation? *Federal Reserve Bank of Chicago WP 2007-02*.

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large n and t panel. *Journal of Business & Economic Statistics 25*(1), 91–96.

Anderson, T. (2003). *An introduction to multivariate statistical analysis*, Volume 2. Wiley New York.

Aron, J. and J. Muellbauer (2013). New methods for forecasting inflation, applied to the us*. *Oxford Bulletin of Economics and Statistics 75*(5), 637–661.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*(1), 137–183.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*(1).

Bai, J. and S. Ng (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*(2), 89–163.

Balke, N. S. and M. A. Wynne (2000). An equilibrium analysis of relative price changes and aggregate inflation. *Journal of Monetary Economics 45*(2), 269–292.

Banerjee, A. and M. Marcellino (2009). Factor-augmented error correction models. In *In: Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 227–254. Oxford University Press.

Beck, G., K. Hubrich, and M. Marcellino (2015). On the importance of sectoral and regional shocks for price-setting. *Journal of Applied Econometrics , forthcomming*.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Bermingham, C. and A. D'Agostino (2014). Understanding and forecasting aggregate and disaggregate price dynamics. *Empirical Economics 46*(2), 765–788.

Beveridge, S. and C. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics 7*(2), 151–174.

Bils, M. and P. J. Klenow (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy 112*(5), 947–985.

Blejer, M. I. and L. Leiderman (1980). On the real effects of inflation and relative-price variability: Some empirical evidence. *The Review of Economics and Statistics 62*, 539–544.

Boivin, J., M. P. Giannoni, and I. Mihov (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *The American Economic Review 99*(1), 350–384.

Boivin, J. and S. Ng (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking 1*, 117–151.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Bollobás, B. and P. Erdös (1976). Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 80, pp. 419–427. Cambridge Univ Press.

Bron, C. and J. Kerbosch (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM 16*(9), 575–577.

Candelon, B., A. Hecq, and W. Verschoor (2005). Measuring common cyclical features during financial turmoil: Evidence of interdependence not contagion. *Journal of International Money and Finance 24*(8), 1317–1334.

Castle, J., J. Doornik, and D. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics 3*(169), 239–246.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting 31*(1), 99–112.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*(2), 239–246.

Castle, J. L. and D. F. Hendry (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting 29*(1-2), 200–214.

Cheung, Y.-W. and K. S. Lai (1993). Finite-sample sizes of johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics 55*(3), 313–328.

Clark, T. (2000). Forecasting an aggregate of cointegrated disaggregates. *Journal of Forecasting 19*(1), 1–21.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics 21*(5), 563–587.

Clements, M. P. and D. F. Hendry (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics 10*(2), 127–146.

Cubadda, G. (1999). Common serial correlation and common business cycles: A cautious note. *Empirical Economics 24*(3), 529–535.

Cubadda, G. (2007). A unifying framework for analysing common cyclical features in cointegrated time series. *Computational Statistics & Data Analysis 52*(2), 896–906.

Cubadda, G. and A. Hecq (2001). On non-contemporaneous short-run co-movements. *Economics Letters 73*(3), 389–397.

D'Agostino, A. and D. Giannone (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics 74*(2), 306–326.

Derényi, I., G. Palla, and T. Vicsek (2005). Clique percolation in random networks. *Physical review letters 94*(16), 160202.

Doornik, J. (2009). Autometrics. In *Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 88–121. Oxford University Press.

Doornik, J., D. Hendry, and B. Nielsen (1998). Inference in cointegrating models: Uk m1 revisited. *Journal of economic surveys 12*(5), 533–572.

Doornik, J., D. Hendry, and F. Pretis (2013). Step-indicator saturation. *University of Oxford. Working Paper* (658).

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics 94*(4), 1014–1024.

Engle, R. and C. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society 55*(2), 251–276.

Engle, R. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics 11*(4), 369–380.

Engle, R. F. and J. V. Issler (1995). Estimating common sectoral cycles. *Journal of Monetary Economics 35*(1), 83–113.

Ericsson, N. (1993). Comment (on engle and kozicki,1993). *Journal of Business & Economic Statistics 11*(4), 380–383.

Escribano, A. and D. Peña (1994). Cointegration and common factors. *Journal of Time Series Analysis 15*(6), 577–586.

Espasa, A. and R. Albacete (2007). Econometric modelling for short term inflation forecasting in the euro area. *Journal of Forecasting 26*(5), 303–316.

Espasa, A. and I. Mayo-Burgos (2013). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting 29*(4), 718–732.

Espasa, A., E. Senra, and R. Albacete (2002). Forecasting inflation in the european monetary union: A disaggregated approach by countries and by sectors. *The European Journal of Finance 8*(4), 402–421.

Fischer, S., R. E. Hall, and J. B. Taylor (1981). Relative shocks, relative price variability, and inflation. *Brookings Papers on Economic Activity 1981*(2), 381–441.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model. *Journal of the American Statistical Association 100*(471).

Franses, P. H. and N. Haldrup (1994). The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business & Economic Statistics 12*(4), 471–478.

Geweke, J. (1977). *The dynamic factor analysis of economic time series models*. Social Systems Research Institute, University of Wisconsin-Madison.

Giacomini, R. and C. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics 118*(1), 7–26.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics 60*(1), 203–233.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*(1), 27–35.

Guerrero, V. M. and D. Peña (2003). Combining multiple time series predictors: a useful inferential procedure. *Journal of Statistical Planning and Inference 116*(1), 249–276.

Hagen, J. v. (1989). Relative commodity prices and cointegration. *Journal of Business & Economic Statistics 7*(4), 497–503.

Hansen, H. and S. Johansen (1999). Some tests for parameter constancy in cointegrated var models. *The Econometrics Journal 2*(2), 306–333.

Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics 16*(4), 388–399.

Harvey, D. I., N. M. Kellard, J. B. Madsen, and M. E. Wohar (2010). The prebisch-singer hypothesis: four centuries of evidence. *The Review of Economics and Statistics 92*(2), 367–377.

Hayek, F. (1945). The use of knowledge in society fa hayek. *American Economic Review 35*(4), 519–30.

Hecq, A., F. Palm, and J. Urbain (2006). Common cyclical features analysis in var models with cointegration. *Journal of Econometrics 132*(1), 117–141.

Hendry, D. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics 135*(1), 399–426.

Hendry, D. and K. Hubrich (2005). Forecasting aggregates by disaggregates. *Computing in Economics and Finance 270*.

Hendry, D. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of business & economic statistics 29*(2), 216–227.

Hendry, D. F. (1995). On the interactions of unit roots and exogeneity. *Econometric reviews 14*(4), 383–419.

Hendry, D. F. and M. Massmann (2007). Co-breaking. *Journal of Business & Economic Statistics 25*(1), 33–51.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005). Ppp strikes back: Aggregation and the real exchange rate. *The Quarterly Journal of Economics 120*(1), 1–43.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control 12*(2-3), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*(6), 1551–1580.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics 52*(3), 389–402.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*, Volume 9. Cambridge University Press.

Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *The Econometrics Journal 3*(2), 216–249.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In *Castle, and Shephard (2009), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 1–36. Oxford University Press.

Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.

Juselius, K. (2006). *The cointegrated VAR model: methodology and applications*. Oxford University Press, USA.

Keynes, J. M. (1924). Tract on monetary reform. *New York: Harcourt, Brace and Co.,*.

Kiviet, J. F. and G. D. Phillips (1992). Exact similar tests for unit roots and cointegration. *Oxford Bulletin of Economics and Statistics 54*(3), 349–367.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics 18*(3), 337–349.

Long Jr, J. B. and C. I. Plosser (1987). Sectoral vs. aggregate shocks in the business cycle. *The American Economic Review 77*(2), 333–336.

Lütkepohl, H. (1984). Linear transformations of vector arma processes. *Journal of Econometrics 26*(3), 283–293.

Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*, Volume 284. Springer Verlag.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2003). Comparison of tests for the cointegrating rank of a var process with a structural shift. *Journal of Econometrics 113*(2), 201–229.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004). Testing for the cointegrating rank of a var process with level shift at unknown time. *Econometrica 72*(2), 647–662.

Lucas, R. E. (1977). Understanding business cycles. In *Carnegie-Rochester conference series on public policy*, Volume 5, pp. 7–29. Elsevier.

Lunnemann, P. and T. Mathä (2004). How persistent is disaggregate inflation? an analysis across eu 15 countries and hicp sub-indices. *Working paper, European Central Bank* (415).

Matula, D. W. (1976). The largest clique size in a random graph. *Department of Computer Science, Southern Methodist University, Technical Report CS 7608*.

Newman, M. (2009). *Networks: an introduction*. OUP Oxford.

Nielsen, B. (2004). On the distribution of likelihood ratio test statistics for cointegration rank. *Econometric reviews 23*(1), 1–23.

Nielsen, B. and H. Nielsen (2008). Properties of estimated characteristic roots. *Univ. of Copenhagen Dept. of Economics Discussion Paper* (08-13).

Nielsen, B. and A. Rahbek (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

Osborn, D. R., A. P. Chui, J. P. Smith, and C. R. Birchenhall (1988). Seasonality and the order of integration for consumption*. *Oxford Bulletin of Economics and Statistics 50*(4), 361–377.

Palla, G., I. Derényi, and T. Vicsek (2007). The critical point of k-clique percolation in the erdős–rényi graph. *Journal of Statistical Physics 128*(1-2), 219–227.

Peña, D. and P. Poncela (2006a). *Dimension reduction in multivariate time series*, pp. 433–458. Springer.

Peña, D. and P. Poncela (2006b). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference 136*(4), 1237–1257.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society 57*, 1361–1401.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Perron, P. and G. Rodríguez (2003). Searching for additive outliers in nonstationary time series*. *Journal of Time Series Analysis 24*(2), 193–220.

Perron, P. and T. J. Vogelsang (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics 10*(3), 301–320.

Pesaran, H. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics 138*(1), 312–355.

Poncela, P. and A. García Ferrer (2014). The effects of disaggregation on forecasting nonstationary time series. *Journal of Forecasting 33*(4), 300–314.

Poncela, P. and E. Ruiz (2012). More is not always better: back to the kalman filter in dynamic factor models. Technical report, WP. Universidad Carlos III, Departamento de Estadística y Econometría.

Quintos, C. E. (1998). Stability tests in error correction models. *Journal of Econometrics 82*(2), 289–315.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies 73*(3), 793–821.

Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegrating rank of a var process with structural shifts. *Journal of Business & Economic Statistics 18*(4), 451–464.

Saikkonen, P. and H. Lutkepohl (2002). Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis 21*(4), 435–456.

Santos, C., D. Hendry, and S. Johansen (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 23*(2), 317–335.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research 1*, 145–168.

Silvester, J. R. (2000). Determinants of block matrices. *The Mathematical Gazette 84*(501), 460–467.

Stock, J. and M. Watson (1988). Testing for common trends. *Journal of the American Statistical Association 83*(404), 1097–1107.

Stock, J. and M. Watson (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Stock, J. and M. Watson (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting, Oxford University Press, Oxford, UK*.

Stock, J., M. Watson, and L. Center (2005). Implications of dynamic factor models for var analysis. *NBER Working paper No. 11467*.

Stock, J. H. and M. W. Watson (2015). Core inflation and trend inflation.

Tiao, G. and R. Tsay (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society. Series B (Methodological) 51*(2), 157–213.

Trenkler, C., P. Saikkonen, and H. Lütkepohl (2007). Testing for the cointegrating rank of a var process with level shift and trend break. *Journal of Time Series Analysis 29*(2), 331–358.

Vahid, F. and R. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics 8*(4), 341–60.

Vahid, F. and R. Engle (1997). Codependent cycles. *Journal of Econometrics 80*(2), 199–221.

Vahid, F. and J. V. Issler (2002). The importance of common cyclical features in var analysis: a monte-carlo study. *Journal of Econometrics 109*(2), 341–363.

Yamamoto, T. (1981). Predictions of multivariate autoregressive-moving average models. *Biometrika 68*(2), 485–492.